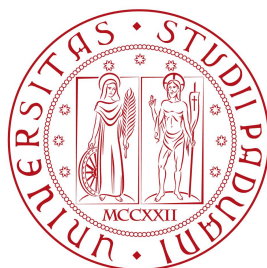


Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Triennale in

Statistica per le Tecnologie e le Scienze



STUDIO STATISTICO SU PAZIENTI AFFETTI DA DISFUNZIONI DELLE PICCOLE VIE AEREE

Relatore: prof.ssa Laura Ventura
Dipartimento di Scienze Statistiche

Laureanda: Emma Lovato
Matricola n. 2008153

Anno Accademico 2022/2023

*A chi è sempre con me
A chi c'è, ma da lontano
A chi non c'è, ma avrebbe voluto esserci
A chi mi ha dedicato il suo tempo.*

Indice

Introduzione	iv
1 Disfunzioni delle piccole vie aeree	1
1.1 Classificazione	1
1.2 Sintomi e cure	2
1.3 Obiettivi e dati	5
2 Analisi esplorativa dei dati	13
2.1 Analisi univariata	13
2.1.1 Variabili qualitative	13
2.1.2 Variabili quantitative	17
2.2 Analisi esplorativa bivariata	24
2.2.1 Variabili qualitative con qualitative . .	25
2.2.2 Variabili quantitative con qualitative .	26
2.2.3 Analisi delle correlazioni	37
2.3 Considerazioni finali	38
3 Regressione Logistica	39
3.1 Modello di regressione logistica	39
3.2 Modellazione	41
3.2.1 Modello finale	42
3.2.2 Interpretazione dei coefficienti	43
3.3 Bontà di adattamento del modello	44
3.4 Considerazioni finali	47

4 Propensity Score Matching	49
4.1 Applicazione al dataset	50
4.2 Analisi dei risultati	54
4.2.1 Analisi bivariata	54
4.2.2 Modello di regressione logistica con il PSM	58
4.2.3 Bontà di adattamento del modello do- po il PSM	60
4.3 Considerazioni finali	62
Risultati e Conclusioni	i
Appendice A	iii
Elenco delle figure	v
Elenco delle tabelle	vii
Bibliografia	ix
Sitografia	xi

Introduzione

La disfunzione delle piccole vie aeree ("SAD" - *Small Airway Dysfunction*) rappresenta un punto cruciale di studio nell'ambito della medicina respiratoria. Queste disfunzioni possono manifestarsi in diverse forme e possono essere correlate a malattie respiratorie come l'asma. Il ruolo della SAD nel controllo e nella gravità dell'asma non è ancora stato ben studiato e spesso è sottovalutato o misconosciuto, non è quindi chiaro se è sempre presente nelle persone che soffrono d'asma e quali siano i test che possono al meglio caratterizzarla.

L'importanza, dunque, di studiare il ruolo della disfunzione delle piccole vie aeree non risiede solamente nella diagnosi precoce e nella gestione delle patologie respiratorie legate ad essa, ma anche nella possibilità di creare e sviluppare terapie mirate alla cura personalizzata di ogni paziente.

Questa tesi di laurea si basa su uno studio clinico condotto su 619 pazienti affetti da asma. Lo scopo è ottenere una miglior comprensione delle cause e verificare l'efficacia e la sicurezza dei trattamenti somministrati con l'obiettivo finale di migliorare la qualità di vita del singolo paziente.

Attraverso questo studio ci si propone di esplorare in profondità la disfunzione delle piccole vie aeree, analizzandone l'eziologia, la patogenesi, la diagnosi e alcune opzioni terapeutiche attualmente disponibili (Vineis *et al.*, 1987).

In particolare, saranno esaminate anche misurazioni di strumenti diagnostici, sia basati sulla spirometria che sull'oscillometria forzata, quali la misurazione del monossido di azoto nelle frazioni di aria espirata (FeNO), il volume espiratorio forzato in un secondo (fEV1p), la *Forced Vital Capacity* cioè il massimo volume espirato nel minor tempo possibile (fVCp) e il rapporto tra le ultime due (fEV1pfVCp), il flusso espiratorio misurato tra il 25% e il 75% della *Forced Vital Capacity* (fEf25-75), la resistenza a 5 Hz (R5), la reattanza a 5 Hz (X5), la resistenza a 5 e 20 Hz (R5R20), l'area di reattanza (AX), la frequenza di risonanza in Hz (Fres) e la differenza tra flusso inspiratorio ed espiratorio (deltaX5). Come trattamento verrà considerata una terapia extrafine con focus sulla sua efficacia nel migliorare la funzione delle piccole vie aeree e nella riduzione dei sintomi respiratori. Il dosaggio di terapia somministrato ai pazienti in mg verrà scelto in base alla gravità dell'asma del paziente determinata dal *Global Initiative for Asthma* (GINA) o dalla Step-terapia.

Infine, questa tesi si propone di sensibilizzare l'importanza della ricerca continua in tale ambito, al fine di sviluppare nuove strategie terapeutiche mirate ai pazienti con disfunzione delle piccole vie aeree in modo da avere una comprensione più approfondita di questo argomento.

Lo schema della tesi è il seguente: nel primo capitolo vengono illustrate, con opportuni riferimenti medici, le disfunzioni delle piccole vie aeree (SAD) per poter comprendere al meglio la successiva analisi preliminare dei dati e le relative elaborazioni effettuate utilizzando il software R. Viene inoltre presentata la struttura del *dataset*: la maggior parte delle variabili è di tipo categoriale e riguardano dati anagrafici e clinici del paziente, vi sono poi alcune variabili numeriche che

rappresentano misurazioni effettuate sui pazienti sottoposti alla terapia studiata.

Nel secondo capitolo vengono effettuate le prime analisi esplorative sui dati, le analisi di correlazione tra le variabili e il calcolo di alcuni test. Vengono inoltre riportati grafici e relativi commenti per una dettagliata analisi del campione di pazienti a cui è stata diagnosticata la SAD.

Nel terzo capitolo si propone un modello di regressione logistica per i dati presi in esame, per individuare quale delle variabili influenzi maggiormente i risultati.

Nel quarto capitolo verrà condotto il *Propensity Score Matching*, una tecnica statistica che permette di migliorare la validità dell'analisi statistica controllando il bias potenziale dovuto a differenze iniziali tra i gruppi di trattamento e di controllo. In particolare si presterà attenzione a come la SAD si comporta nei confronti della popolazione con meno di 65 anni rispetto a chi ne ha più di 65.

Nelle Conclusioni vengono riepilogati i vari metodi di studio applicati per i dati in esame e quindi riassunti in breve i risultati ottenuti dalle varie elaborazioni.

Capitolo 1

Disfunzioni delle piccole vie aeree

Le disfunzioni delle piccole vie aeree (SAD) sono uno spettro di patologie infiammatorie o fibrotiche caratterizzate da una limitazione del flusso d'aria nelle piccole vie aeree. Queste condizioni possono causare difficoltà nella respirazione e influire sulla funzione polmonare. Le piccole vie aeree, bronchioli o alveoli, sono prive di strutture cartilaginee e hanno un diametro inferiore ai 2 mm (Figura 1.1).

1.1 Classificazione

Alcune delle disfunzioni delle piccole vie aeree più comuni includono:

- **Bronchiolite:** Si tratta di un'infezione virale che coinvolge la parte finale dei bronchi, i cosiddetti bronchioli terminali, provocandone l'infiammazione e la congestione. È molto comune nei bambini e può causare tosse, respiro sibilante e difficoltà respiratorie.
- **Bronchiectasie:** sono una malattia respiratoria cronica caratterizzata da un danno permanente delle pareti dei

bronchioli, che si allargano e perdono la loro elasticità, con una conseguente maggior produzione di muco e infezioni ricorrenti delle vie aeree.

- **Asma:** è una malattia infiammatoria che colpisce in particolare i bronchi, causandone un restringimento temporaneo, associato ad un eccesso di produzione di muco. I sintomi includono tosse, respiro sibilante, oppressione toracica e difficoltà respiratorie.
- **Fibrosi cistica:** è una malattia genetica che colpisce le ghiandole esocrine, volte alla produzione di muco e sudore. La fibrosi cistica è caratterizzata dalla produzione di muco denso e appiccicoso che si deposita bloccando le vie respiratorie, portando a infezioni respiratorie ricorrenti e danni polmonari progressivi.
- **Broncopneumopatia cronica ostruttiva (BPCO):** La BPCO è una malattia polmonare cronica ostruttiva che coinvolge polmoni, bronchi e bronchioli. Causa una riduzione del flusso d'aria e sintomi come tosse cronica, produzione di muco e provoca difficoltà respiratorie.

1.2 Sintomi e cure

I sintomi delle disfunzioni delle piccole vie aeree (SAD) possono variare a seconda delle condizioni patologiche sopracitate; tuttavia, vi sono alcuni sintomi ricorrenti che possono essere associati alla SAD. I più comuni possono essere:

- **Tosse:** può essere il sintomo di diversi disturbi o malattie, sia che si tratti di una forma secca, sia che si tratti di tosse grassa, cioè, associata all'espettorazione di muco. Se la

tosse persiste per più di 3 settimane viene definita cronica e potrebbe essere l'effetto di problemi più seri, come un enfisema o bronchite cronica.

- **Respiro sibilante:** è un respiro associato a un suono acuto, simile a un fischio, che può segnalare la presenza di problemi respiratori ed è causato dalla restrizione del flusso d'aria attraverso le piccole vie aeree, più precisamente i bronchi
- **Difficoltà respiratorie:** Questo sintomo può variare da lieve a grave. Le persone con SAD possono sperimentare sensazioni di mancanza d'aria, difficoltà a respirare profondamente e sensazioni di oppressione toracica.
- **Aumento della produzione di muco:** In alcune condizioni la SAD può portare il paziente a secernere muco in quantità considerevoli, che può causare congestione e tosse.
- **Asma da sforzo:** diversamente dall'asma bronchiale, l'asma da sforzo si presenta solo durante l'attività sportiva
- **Risvegli notturni a causa dell'asma:** è l'insorgenza di sintomi respiratori durante la notte o nelle prime ore del mattino. Le cause notturne delle manifestazioni d'asma sono molteplici, tra cui:
 1. La posizione supina
 2. Maggiore esposizione agli allergeni
 3. Raffreddamento delle vie respiratorie e umidità dell'aria, dovute eventualmente all'aria condizionata
 4. Fattori neurormonali

- Ricorrenti riacutizzazioni: Le persone con SAD possono sperimentare episodi di peggioramento dei sintomi o di ripresa dell'andamento acuto della patologia, noti come esacerbazioni o riacutizzazioni, che richiedono spesso cure mediche aggiuntive.

È importante sottolineare che questi sintomi possono sovrapporsi tra le diverse condizioni della SAD e portare quindi i pazienti a condizioni fisiche gravi.

Il trattamento delle disfunzioni delle piccole vie aeree dipende dunque dalla causa e dalla gravità dei sintomi. Una delle forme più efficaci di trattamento dei sintomi causati dalle disfunzioni delle piccole vie aeree attualmente in commercio è la terapia extrafine, che coinvolge l'uso di farmaci sotto forma di aerosol o polveri finissime inalate attraverso un dispositivo specifico. Il principale vantaggio di questo tipo di terapia rispetto ai farmaci convenzionali per la cura dell'asma è la dimensione delle particelle estremamente piccole. Queste particelle extrafini sono progettate per penetrare più profondamente nelle vie respiratorie, raggiungendo le piccole vie aeree, inclusi bronchioli e alveoli (Humanitas, 2023).

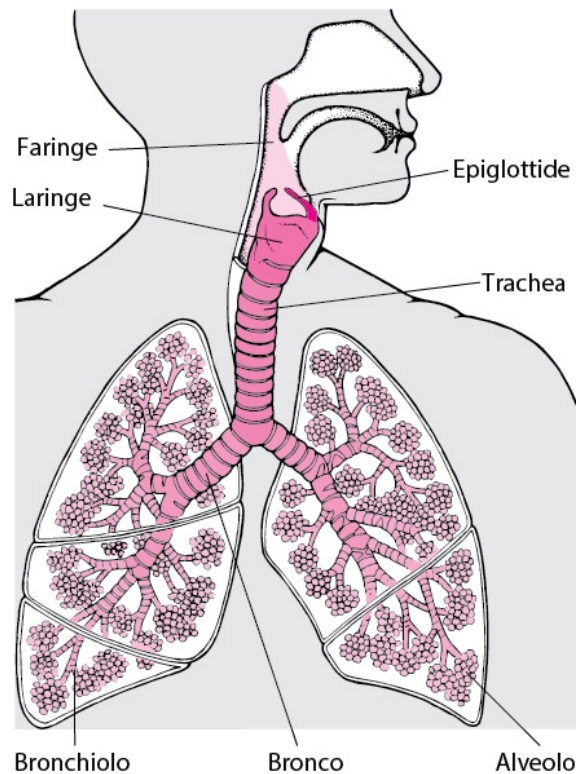


Figura 1.1: Panoramica sull'apparato respiratorio.

1.3 Obiettivi e dati

Il *dataset* a disposizione contiene informazioni su 619 pazienti che possono essere affetti da disfunzioni delle piccole vie aeree e curati con terapia extrafine, un totale di 30 variabili compongono le informazioni del *dataset*.

Obiettivo principale dello studio è analizzare i fattori che influenzano la presenza delle disfunzioni delle piccole vie aeree (SAD) al fine di identificare i principali determinanti di questa condizione.

In particolare, si vogliono studiare le variabili indipendenti quali l'età, il sesso, la durata dell'asma nei pazienti, la predisposizione ereditaria alle malattie allergiche, i livelli di gravità dell'asma determinati dal *Global Initiative for Asthma*, la ripresa dell'andamento acuto della patologia, l'acc-

cesso al pronto soccorso da parte del paziente, il ricovero del paziente, la quantità di eosinofili contenuti nei bronchi, la strategia di gestione dell'asma dettata dalla Step-terapia, la storia tabagica del paziente, l'indice di massa corporea, i risvegli notturni causati dall'asma, l'asma da sforzo per comprendere le loro associazioni con la SAD.

Si tiene conto inoltre della terapia extrafine somministrata ai pazienti e del dosaggio in mg di questa, degli strumenti diagnostici, quali la misurazione di monossido di azoto nell'aria espirata, il volume espiratorio forzato nel primo secondo, la *Forced Vital Capacity*, il rapporto tra queste ultime due quantità, il flusso espiratorio tra il 25% e il 75% della *Forced Vital Capacity*, la difficoltà di un gas a muoversi all'interno dei polmoni a 5 e a 20 Hz, la capacità dei polmoni a contenere gas, l'area di reattanza, la frequenza di risonanza in Hz e la differenza tra flusso inspiratorio ed espiratorio, utilizzati per valutare la natura della patologia nel paziente.

Si suddivide il *dataset* in base alla natura delle variabili.

Le variabili di natura qualitativa sono:

- SAD, indica la presenza (1) o l'assenza (0) di disfunzioni delle piccole vie aeree nel paziente;
- SESSO, indica se il paziente è di genere maschile (M) o femminile (F);
- OVER65, è stato inserito un cut-off nella variabile *eta* in modo da suddividere il campione in due parti, ovvero chi ha più di 65 anni (1) e chi meno di 65 anni (0);
- ATOPIA, è la predisposizione ereditaria del paziente alle malattie allergiche;

- GINA (*Global Initiative for Asthma*), è un'organizzazione di linee guida mediche che lavora con i funzionari della sanità pubblica e gli operatori sanitari a livello globale per ridurre la prevalenza, la morbilità e la mortalità dell'asma. Queste linee guida si basano su un regolatore farmaco-dose, l'asma risulta quindi essere classificato come ben controllabile (WC, assenza di sintomi), parzialmente controllabile (PC, 1 o 2 sintomi), e incontrollabile (U, più di 3 sintomi);
- GINA1, questa variabile accorpa i due livelli PC e U di GINA;
- RIACUTIZZAZIONI, indica se il paziente ha (1) o meno (0) una ripresa dell'andamento acuto del processo morboso;
- ACCESSIPS, ovvero se il paziente ha avuto accesso (1) o meno (0) al Pronto Soccorso durante la durata dello studio;
- RICOVERI, indica se il paziente è stato ricoverato (1) o no (0) durante il periodo di studio per sintomi causati dall'asma;
- STEP-TERAPIA; è una strategia di gestione dell'asma che coinvolge l'uso progressivo di farmaci in base alla gravità dei sintomi e alla risposta del paziente. È una variabile categoriale divisa in 5 livelli (da 1 a 5) che gradualmente indicano lo stato di gravità dell'asma nel paziente;
- FUMO, indica se il paziente fuma (1) o meno (0);

- RISVEGLI, ovvero se il paziente durante almeno una notte nel periodo di osservazione si è svegliato (1) o no (0) per sintomi causati dall'asma;
- EIA (exercise-induced asthma), indica la presenza (1) o assenza (0) di asma da sforzo nel paziente;
- TERAPIAEXTRAFINE, indica se il paziente è stato trattato (1) o meno (0) con terapia inalatoria extrafine, che si distingue dalle altre terapie per la dimensione delle particelle estremamente piccole che sono progettate per penetrare più profondamente nelle vie respiratorie, raggiungendo le piccole vie aeree, inclusi bronchioli e alveoli.

Le variabili di natura quantitativa sono:

- ETA, rappresenta l'età del paziente al momento della diagnosi della SAD;
- DURATSMA, rappresenta la durata in mesi dell'asma nel paziente;
- EOSINOFILI, indica il numero di globuli bianchi (leucociti), coinvolti nelle reazioni allergiche e nella difesa contro le infestazioni parassitarie, presenti nei bronchi del paziente al momento della diagnosi;
- BMI, indice di massa corporea del paziente;
- DOSAGGIO ICS, rappresenta il dosaggio di terapia somministrato al paziente in mg.

Alcune variabili quantitative rappresentano i risultati di test diagnostici misurati attraverso la spirometria:

- FENO, Si tratta di un esame non invasivo utilizzato per le malattie respiratorie allergiche che misura i livelli di ossido nitrico (NO, o monossido di azoto) presente nell'aria espirata dal paziente. La presenza di questo prodotto chimico nell'aria espirata dalla bocca o dal naso del paziente è correlata all'infiammazione riscontrabile nell'albero respiratorio e, in particolar modo, a quella quota della stessa prodotta dagli eosinofili, presenti nei bronchi, la cui presenza spesso si associa alla patologia asmatica allergica (Ariano, 2023);
- FEV1P, è un importante indice di pervietà delle vie aeree, indica il volume espiratorio massimo in 1 secondo, il risultato della spirometria è da considerarsi normale quando i valori di $fEV1p$ sono compresi tra 80% e 120% del valore medio; valori bassi del $fEV1p$ sono indicativi di un problema ostruttivo (Ariano, 2023);
- FVCP (*Forced Vital Capacity*), indica il volume totale di aria che si riesce ad espellere in espirazione forzata, dopo aver eseguito un'inspirazione massimale;
- FEV1FVCP, è il rapporto tra $fEV1p$ e $fVCp$, indica la percentuale della capacità vitale forzata che viene espirata al primo secondo. Per convenzione un rapporto < 70 indica ostruzione;
- FEF25-75, indica il flusso espiratorio tra il 25% e il 75% della *Forced Vital Capacity*.

La spirometria consiste nella misurazione della quantità di aria che il polmone utilizza nella respirazione e nella misura dei flussi dell'aria durante la respirazione forzata confrontando i risultati con uno standard noto (Humanitas, 2023).

Alcune variabili quantitative rappresentano i risultati dati da test diagnostici misurati attraverso l'oscillometria forzata:

- R5, indica la resistenza a 5 Hz, cioè la difficoltà del gas a muoversi all'interno del polmone a causa della resistenza che incontra nelle vie aeree centrali e periferiche;
- X5, indica la reattanza a 5 Hz, variabile che stima la capacità del polmone a contenere gas. Viene misurata con l'oscillometria forzata;
- R5R20, indica la resistenza a 5 e a 20 Hz;
- AX, rappresenta l'area di reattanza sotto la curva o reattanza totale a tutte le frequenze tra i 5 Hz e F_{res} ;
- FRES, ovvero la frequenza di risonanza. Le frequenze intermedie a cui la reattanza totale è 0. La reattanza a 5 Hz, insieme a F_{res} e AX, viene considerata come un indice di anomalie delle vie aeree periferiche e ne riflettono tutti i cambiamenti nei diversi gradi di ostruzione (Goldman *et al.*, 2005);
- DELTAX5, rappresenta la differenza tra flusso inspiratorio e espiratorio a 5 Hz.

L'oscillometria forzata è una tecnica semplice, non invasiva per calcolare l'impedenza del sistema toraco-polmonare (impedenza = resistenza + reattanza). Per impedenza s'intende la forza di opposizione di un circuito al passaggio di un impulso (Boccaccino, 2005).

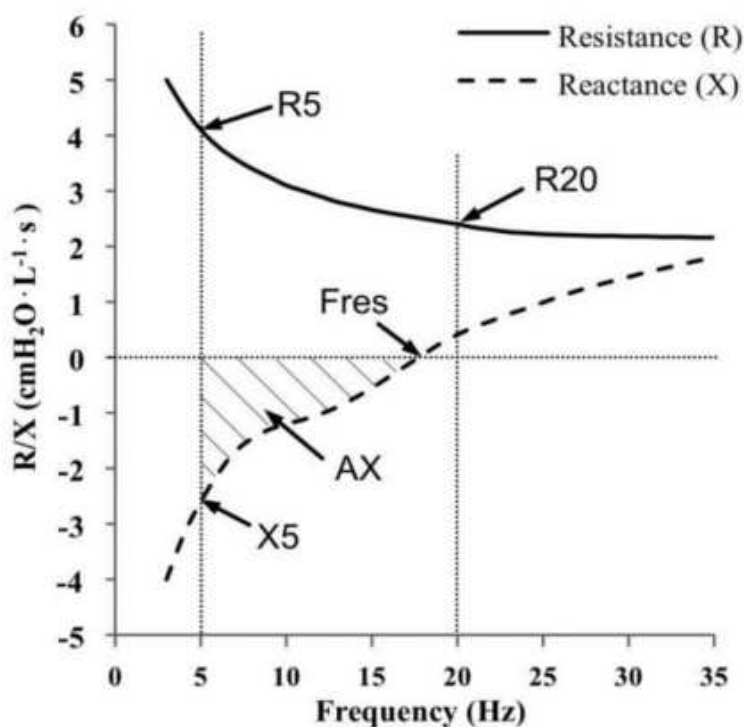


Figura 1.2: Rappresentazione schematica degli indici IOS rispetto alle frequenze di oscillazione: R5, R20, X5, AX e fres.

Dopo un'analisi esplorativa dei dati e un'illustrazione generale delle variabili del *dataset* (Capitolo 2), si procede con l'analisi della variabile d'interesse (Capitolo 3), che rappresenta la presenza o meno della disfunzione delle piccole vie aeree nei pazienti attraverso un modello di regressione logistica e l'analisi dei dati appaiati attraverso il *Propensity Score Matching* (Capitolo 4).

Si precisa che durante tutta l'analisi si lavora con un livello di significatività del 5%.

Nel *dataset* originario sono stati riscontrati valori mancanti (1.38% di NA) che sono stati successivamente imputati utilizzando la mediana al fine di effettuare l'analisi.

Dalle 619 unità statistiche iniziali viene inoltre rimossa l'unità 569, dal momento che presenta un valore mancante per la variabile *EIA*, il *dataset* si compone così di 618 unità.

Capitolo 2

Analisi esplorativa dei dati

In questo capitolo si effettua una prima analisi esplorativa di tutte le variabili osservate per i 618 pazienti sottoposti a terapia extrafine o meno. Ai metodi grafici, quali boxplot, si affiancano opportuni test.

Di seguito, in particolare, verranno considerati i test non parametrici basati sui ranghi (Piccolo, 1998), dal momento che il test di Shapiro-Wilk, per la verifica della normalità della distribuzione, porta a rifiutare l'ipotesi nulla per tutte le variabili presenti nel *dataset* (p-value < 0.0001).

2.1 Analisi univariata

2.1.1 Variabili qualitative

Di seguito, nella Tabella 2.1, sono riportate le frequenze assolute e percentuali delle variabili qualitative.

Il campione è composto da 404 pazienti affetti da disfunzione delle piccole vie aeree (65.3%) e da 215 pazienti che non ne soffrono (34.7%).

Per quanto riguarda il sesso, il campione si divide abbastanza omogeneamente tra maschi e femmine (55.9% F e 44.1% M).

Se si suddivide il campione in esame inserendo un cut-off all'età di sessantacinque anni, i pazienti con più di sessantacinque anni risultano essere 1/4 di quelli con meno di sessantacinque anni.

Circa il 61% del campione è predisposto ereditariamente a malattie allergiche.

Se si suddividono i pazienti seguendo le linee guida introdotte dal *Global Initiative for Asthma*, risulta che la maggior parte di essi soffre d'asma parzialmente controllabile (40% PC), circa il 34% soffre di asma ben controllabile e solamente il 26% soffre di asma incontrollabile, accorpendo invece i due livelli PC e U di *GINA* si ottiene una percentuale più numerosa di pazienti con asma classificato come più grave (65.4%).

La maggior parte dei pazienti non ha mai avuto riacutizzazioni (62.5%), o non è mai andata al pronto soccorso (89.7%) o stata ricoverata (96.9%) durante tutto il periodo dello studio.

Se si partiziona il campione secondo la strategia basata sulla Step-terapia, i pazienti che presentano asma intermittente sono solamente il 2.9%, quelli con asma lieve persistente sono il 17.1%, quelli con asma moderato persistente sono la maggioranza (52.8%), chi ha asma grave persistente invece fa parte del 24.7% del campione e il rimanente 2.4% ha asma molto grave persistente.

Solamente il 37% dei pazienti è stato curato con terapia inalatoria extrafine, per il restante 63% non è dato a sapersi se venga trattato con un altro tipo di terapia o non venga

trattato con alcuna terapia.

Il 33.6% della popolazione oggetto di studio fuma, mentre il 66.4% no.

Una buona parte dei pazienti (44.3%) soffre di risvegli notturni causati dall'asma.

Poco più della metà del campione (52.8%) soffre di asma da sforzo quando pratica esercizio fisico.

VARIABILE	CATEGORIE	FREQ. ASS.	FREQ. %
SAD	Si	404	65.3
	No	215	34.7
SESSO	Femmine	346	55.9
	Maschi	273	44.103
OVER65	Si	162	26.2
	No	457	73.8
ATOPIA	Si	380	61.4
	No	239	38.6
GINA	WC	214	34.6
	PC	247	39.9
	U	158	25.5
GINA1	WC	214	34.6
	PC-U	405	65.4
RIACUTIZZAZIONI	Si	232	37.5
	No	387	62.5
ACCESSIPS	Si	64	10.3
	No	555	89.7
RICOVERI	Si	19	03.1
	No	600	96.9
STEP TERAPIA	Intermittente	18	2.9
	Lieve	106	17.1
	Moderato	327	52.8
	Grave	153	24.7
	Molto grave	15	2.4
TERAPIA EXTRA FINE	Si	229	36.9
	No	390	63.1
FUMO	Si	208	33.6
	No	411	66.4
RISVEGLI	Si	262	42.3
	No	357	57.7
EIA	Si	327	52.8
	No	291	47.2

Tabella 2.1: Frequenze assolute e percentuali delle variabili qualitative.

2.1.2 Variabili quantitative

Nella Tabella 2.2 si riportando gli indicatori di sintesi delle variabili quantitative.

VARIABILE	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Sd
ETA	18.000	37.000	51.000	51.030	66.000	94.000	18.209
DURATSMA	1.000	6.000	15.000	17.100	22.000	251.000	16.194
FENO	2.000	13.000	23.000	31.140	43.000	172.000	24.598
EOSINOFILI	110.000	230.000	315.000	341.700	388.000	1700.000	161.612
FEV1P	35.000	79.000	93.000	91.120	104.000	145.000	19.436
FVCP	48.000	89.000	100.000	101.800	111.000	1117.000	44.156
FEV1PFVCP	36.000	67.450	74.100	73.490	80.150	102.000	9.819
FEF25-75P	10.000	46.000	68.000	70.410	89.000	185.000	31.528
R5	0.150	0.350	0.450	0.486	0.570	1.490	0.196
X5	-0.700	-0.220	-0.150	-0.175	-0.105	0.080	0.102
R5R20	0.000	0.050	0.110	0.132	0.185	0.640	0.107
AX	0.000	0.430	1.130	1.503	2.100	10.180	1.417
FRES	0.431	14.285	21.065	20.506	25.690	54.860	7.787
DELTA5	-0.080	0.000	0.020	0.082	0.110	1.200	0.142
BMI	10.000	22.000	24.000	25.130	28.000	49.000	5.171
DOSAGGIO ICS	100.000	400.000	750.000	854.100	1000.000	4000.000	571.694

Tabella 2.2: Indicatori di sintesi della distribuzione delle variabili quantitative.

Variabili che comprendono caratteristiche generali del paziente

Di seguito vengono esplorate alcune variabili che includono alcune informazioni di carattere generale del paziente, quali l'età, la durata dell'asma in mesi, la quantità di eosinofili misurata nei bronchi, l'indice di massa corporea e il dosaggio di terapia somministrato in mg (Figura 2.1).

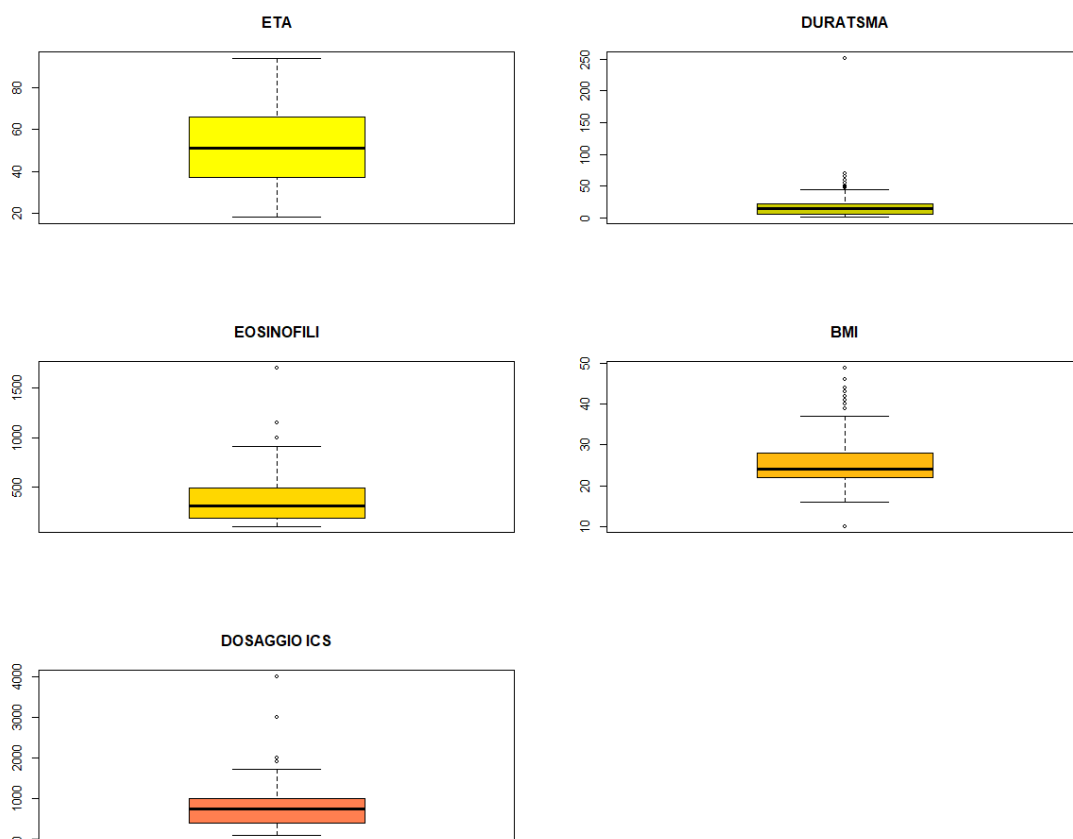


Figura 2.1: Boxplot delle variabili generali del paziente.

La variabile *eta* rappresenta l'età in anni della popolazione presa in esame. L'età considerata al momento della diagnosi è compresa tra i 18 e i 94 anni, con una media di 51 anni circa e una deviazione standard pari a 18.2. Dal grafico riportato in Figura 2.1 si può notare che il 50% dei pazienti ha tra i 37 e 66 anni.

La variabile *duratsma* si riferisce alla durata in mesi dell'asma del paziente. La durata dell'asma è compresa tra 1 e 251 mesi, con una media di circa 17 mesi e una deviazione standard pari a 16.2. Dal grafico riportato in Figura 2.1 si può vedere che la durata dell'asma nel 50% dei pazienti va dai 6 ai 22 mesi.

La variabile *Eosinofili* quantifica il numero di leucociti

coinvolti nelle reazioni allergiche presenti nei bronchi del paziente. La loro presenza spesso si associa alla patologia asmatica allergica. Gli eosinofili nei pazienti sono compresi tra 110 e 1700, con una media di 341.7 e una deviazione standard pari a 161.6. Il boxplot in Figura 2.1 mostra come la metà della distribuzione abbia tra i 230.000 e i 388.000 eosinofili nei bronchi.

L'Indice di Massa Corporea (BmI) è calcolato dividendo il peso in kilogrammi per il quadrato dell'altezza in metri. Per quanto concerne l'indice di massa corporea si osserva che va da un minimo di 10 (un BmI di 10 corrisponde a un peso di circa 28,9 chilogrammi con un'altezza di 1.70 metri) a un massimo di 49 (un BMI di 49 corrisponderebbe a un peso di circa 140.61 chilogrammi con un'altezza di 1.70 metri). In media si ha un valore di BmI pari a 25.13 e una deviazione standard pari a 5.171.

La variabile *Dosaggio ICS* quantifica il dosaggio in mg di terapia extrafine somministrata ai pazienti dello studio. Il dosaggio minimo è pari a 100 mg, il dosaggio massimo è 4000, si ha una media di 854.1 con una deviazione standard pari a 571.7.

Variabili relative alla spirometria

Si riportano ora le variabili relative ai risultati ottenuti attraverso il test spirometrico nei pazienti, ovvero la quantità di ossido nitrico nell'aria espirata, l'indice di pervietà delle vie aeree $fEV1$, la *Forced Vital Capacity*, il rapporto tra queste due ultime quantità e il flusso di aria espirata tra il 25% e il 75% della *Forced Vital Capacity* (Figura 2.2).

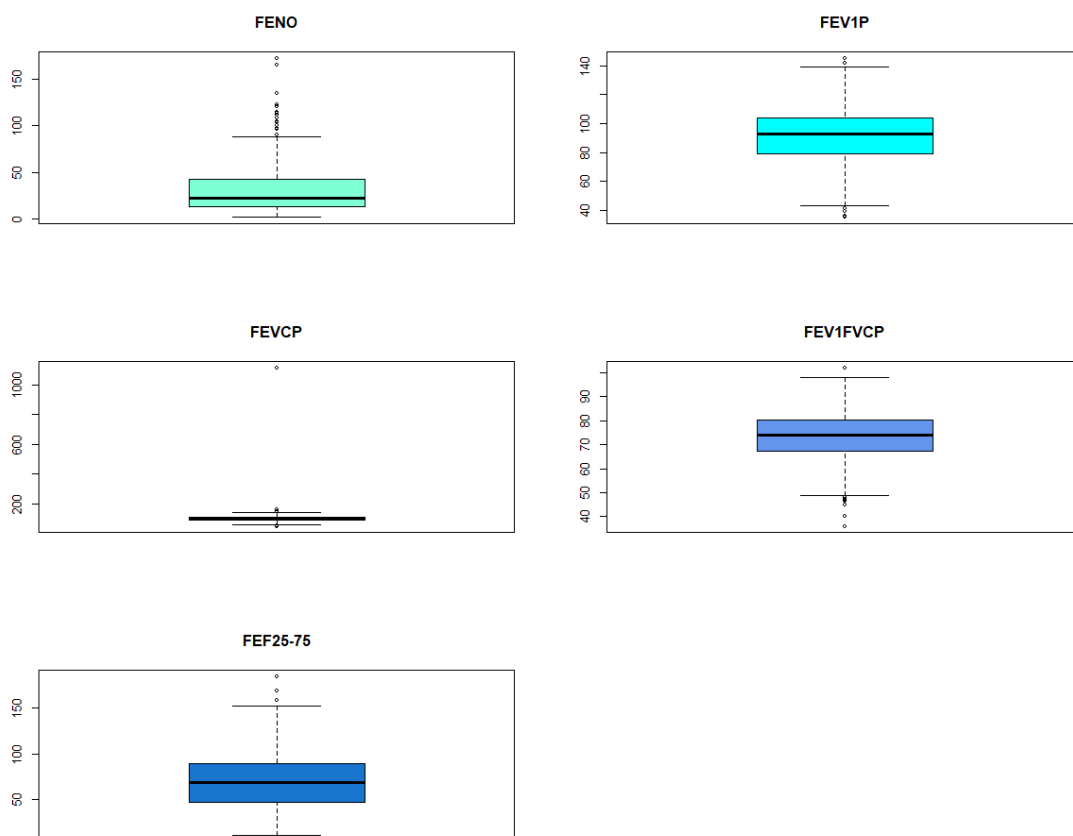


Figura 2.2: Boxplot delle variabili relative ai risultati del test diagnostico spirometrico.

L'ossido nitrico (NO) è un marker di infiammazione per cui la sua concentrazione nell'aria espirata permette di valutare il grado di infiammazione delle vie aeree. Si tratta di un test non invasivo che può rappresentare un valido strumento alternativo a tecniche invasive per la diagnosi e il monitoraggio dell'asma bronchiale (Ariano, 2023). L'ossido nitrico espirato presente nei pazienti è compreso tra 2 e 172, con una media di 31.140 e una deviazione standard pari a 24.6. Dal boxplot riportato in Figura 2.2 si nota che la metà della popolazione presa in esame ha ossido nitrico compreso tra 13 e 43 nell'aria espirata.

La variabile $fEV1p$ rappresenta un importante indice di

pervietà delle vie aeree. Valori bassi di $fEV1p$ indicano un problema ostruttivo nelle piccole vie aeree del paziente. Valori bassi di $fEV1p$ indicano un problema ostruttivo. L'indice $fEV1p$ va da un minimo di 35 ad un valore massimo pari a 145, mediamente si ha 91.1 con una deviazione standard di 19.4. Il boxplot in Figura 2.2 mostra come la metà dei pazienti abbia un valore che va da 79 a 104 di $fEV1p$.

La variabile $fVCp$ indica la *Forced Vital Capacity* dei pazienti, va da un minimo di 48 a un massimo di 1117 di massimo volume espirato nel minor tempo possibile. In media si ha 101.8 di volume con una deviazione standard pari a 44.2.

Il rapporto fisso $fEV1p/fVCp < 0.70$ frequentemente utilizzato come limite inferiore di normalità genera una sottostima della condizione patologica (falsi negativi) nei soggetti di età < 50 anni e un eccesso di diagnosi (falsi positivi) nei soggetti di età superiore a 50 anni. Sarebbe preferibile utilizzare, come limite inferiore di normalità, il 95° percentile del valore di riferimento del rapporto FEV1/VC, comunemente conosciuto come indice di Tiffeneau (con VC = capacità vitale) (Lombardo *et al.*, 2019). Avendo a disposizione solamente questo indice però, si nota come in media i pazienti superino il rapporto fisso $fEV1p/fVCp$ (73.49) con una deviazione standard pari a 9.8 e che la metà dei pazienti abbia un valore di $fEV1p/fVCp$ che va da 67.5 a 80.2, indice che la maggior parte dei pazienti non presenta ostruzione, sebbene di poco, delle piccole vie aeree. Si tiene conto che, dal momento che la maggior parte della popolazione (73.8%) ha età inferiore a sessantacinque anni, potrebbe quindi verificarsi una sottostima della condizione patologica dei pazienti.

La variabile $fEf25-75p$ indica il flusso espiratorio tra il 25% e il 75% della *Forced Vital Capacity*. Il $fEf25-75p$, essendo

un indice molto variabile, va preso in considerazione solo se aumenta oltre il 45% e mai da solo, ma considerando sempre la storia clinica del soggetto, valutandolo nel tempo con altre spirometrie e con altri test di funzionalità respiratoria e assicurandosi che i valori della Forced Vital Capacity siano normali (Boccaccino, 2018). I valori di fEf_{25-75p} vanno da un minimo di 10.000 ad un massimo di 185, in media si ha 70.4 con una deviazione standard pari a 31.5 di flusso espiratorio tra i 25% e il 75%.

Variabili relative all'oscillometria forzata

Di seguito si riportano le variabili indicanti i risultati ottenuti dal test basato sull'oscillometria forzata nei pazienti, ovvero le due variabili per l'impedenza polmonare, l'indice di resistenza delle vie aeree periferiche, la reattanza totale, la frequenza di risonanza in Hz e la differenza tra flusso inspiratorio ed espiratorio di $X5$ (Figura 2.3).

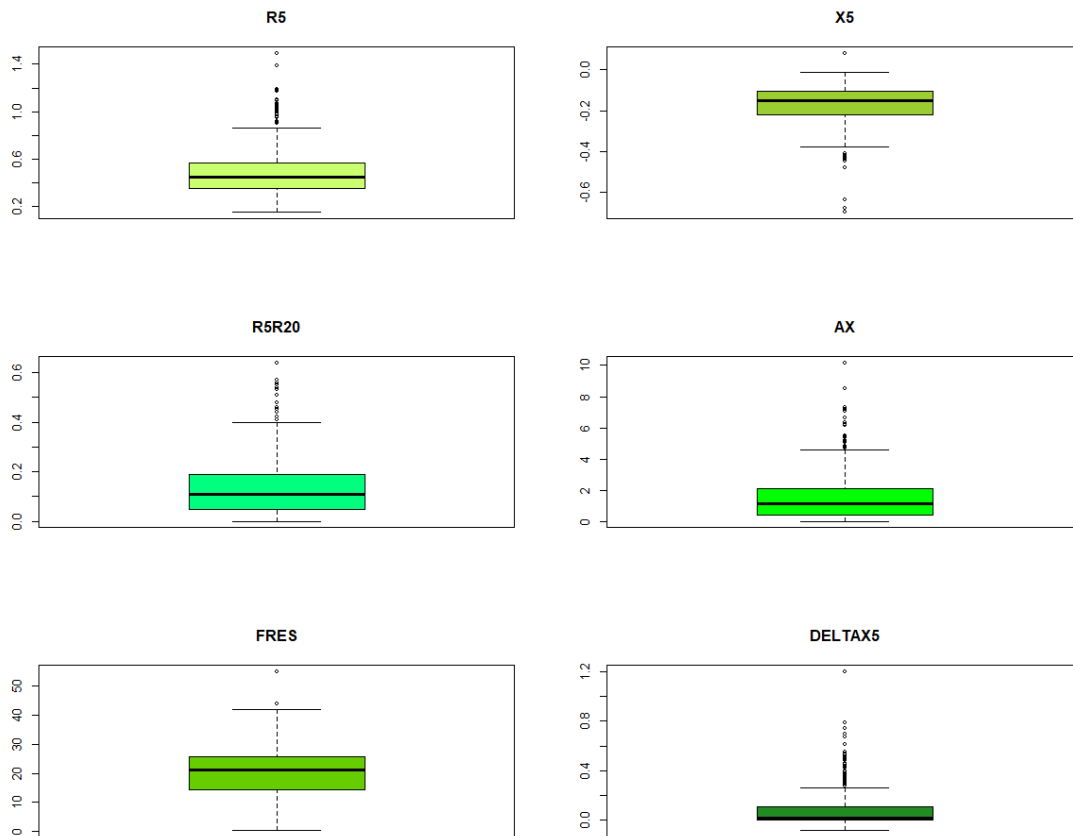


Figura 2.3: Boxplot delle variabili relative ai risultati del test diagnostico basato sull'oscillometria forzata.

La variabile $R5$ indica la resistenza a 5 Hz e si riferisce alla difficoltà del gas a muoversi all'interno del polmone a causa della resistenza che incontra nelle vie aeree centrali e periferiche. La resistenza va da un minimo di 0.15 ad un massimo di 1.5, con media pari a 0.49 e deviazione standard 0.196.

La variabile $X5$ indica la reattanza a 5 Hz e stima la capacità del polmone a contenere gas, assieme alla resistenza costituisce l'impedenza polmonare. Va da un minimo di -0.7 Hz ad un massimo di 0.08 Hz nei pazienti dello studio, con una media pari a -0.18 e una deviazione standard uguale a 0.102.

Basse frequenze di oscillazione ($<15\text{Hz}$) possono essere trasmesse più distalmente rispetto alle alte frequenze (124), R a 5Hz viene considerata dunque un indice della resistenza totale mentre R a 20 Hz come indice delle vie prossimali. La differenza nella resistenza tra 5 e 20 Hz è considerata indice della resistenza solo delle vie aeree periferiche (Grimby *et al.*, 1968). In questo studio l'indice $R5R20$ va da un minimo pari a 0.0 ad un massimo pari a 0.64, in media si ha 0.132 con una deviazione standard di 0.107.

La variabile AX rappresenta l'area di reattanza o reattanza totale, ha un valore minimo pari a 0 e un valore massimo pari a 10.2, in media si ha 1.5 con una deviazione standard di 1.4. Dal boxplot in Figura 2.3 si nota che il 50% dei pazienti è compresa tra i valori 0.4 e 2.1 di AX .

La variabile $Fres$ rappresenta la frequenza di risonanza in Hz, va da un minimo pari a 0.4 a un massimo pari a 54.9, con una media di 20.5 e la deviazione standard di 7.8.

La variabile $deltaX5$ si riferisce alla differenza tra flusso inspiratorio ed espiratorio di X5. $deltaX5$ va da un minimo di -0.08 ad un massimo di 1.2, con media pari a 0.08 e deviazione standard 0.14.

2.2 Analisi esplorativa bivariata

Di seguito si riporta l'analisi esplorativa bivariata, in particolare vengono inserite solamente le relazioni tra variabili risultate significative dai test di verifica di ipotesi, quali il chi-quadro e i test non parametrici di Mann-Whitney e Kruskal-Wallis basati sui ranghi.

Si esegue inoltre un'analisi delle correlazioni delle variabili quantitative in modo da determinare se esiste un legame tra

tali variabili e quantificarne la relazione.

2.2.1 Variabili qualitative con qualitative

Nella Tabella 2.3 si riportano gradi di libertà, valore della statistica test e p-value del test chi-quadro, in particolare vengono riportate le relazioni risultate significative.

COPPIE DI VARIABILI	G.L.	X^2	P-VALUE
SESSO e SAD	1	12.360	0.0004
OVER65 e SAD	1	26.428	< 0.0001
ATOPIA e SAD	1	16.622	< 0.0001
GINA e SAD	2	188.020	< 0.0001
GINA1 e SAD	1	159.580	< 0.0001
RIACUTIZZAZIONI e SAD	1	82.494	< 0.0001
ACCESSIPS e SAD	1	26.967	< 0.0001
RICOVERI e SAD	1	8.911	0.003
STEPTEAPIA e SAD	4	15.494	0.004
TERAPIAEXTRAFINE e SAD	1	82.546	< 0.0001
FUMO e SAD	1	32.189	< 0.0001
RISVEGLI e SAD	1	117.720	< 0.0001
EIA e SAD	1	205.980	< 0.0001
FUMO e OVER65	1	9.571	0.002
FUMO e EIA	1	27.965	< 0.0001
FUMO e RIACUTIZZAZIONI	1	17.126	< 0.0001
RIACUTIZZAZIONI e RICOVERI	1	30.003	< 0.0001
RIACUTIZZAZIONI e ACCESSIPS	1	104.660	< 0.0001
RIACUTIZZAZIONI e ATOPIA	1	18.068	< 0.0001
ACCESSIPS e GINA	2	159.190	< 0.0001
ACCESSIPS e STEPTEAPIA	4	16.822	0.002

Tabella 2.3: Risultati dei test X^2 per l'indipendenza delle variabili qualitative.

Nella Tabella 2.3 si osservano i risultati del test del chi-quadro che valuta le relazioni tra diverse variabili qualitative e la variabile critica delle disfunzioni delle piccole vie aeree (SAD).

È evidente che tutte le relazioni presentate nella tabella sono statisticamente significative, confermando l'importanza di queste variabili nel contesto delle disfunzioni delle piccole vie aeree. Questi risultati suggeriscono un legame sostanziale tra le variabili qualitative in questione e la presenza di SAD.

Vi sono inoltre alcune interazioni tra variabili qualitative risultate significative, in particolare ci si aspetta che nella formulazione del modello di regressione logistica nel Capitolo 3 risultino significative le relazioni tra fumare e avere più o meno di sessantacinque anni, fumare e l'asma da sforzo fisico, fumare e la ripresa dell'andamento acuto dell'asma, le riacutizzazioni e il ricovero del paziente, le riacutizzazioni e il recarsi al pronto soccorso, le riacutizzazioni e l'aver una predisposizione ereditaria alle malattie allergiche, gli accessi a pronto soccorso e la classificazione di asma nel paziente dato dal *Global Initiative for Asthma* o dalla step-terapia.

2.2.2 Variabili quantitative con qualitative

Si verifica inizialmente la normalità nei gruppi utilizzando il test di Shapiro-Wilk, ne risulta che per ogni relazione tra variabili la normalità non è mai rispettata per almeno un gruppo, ne consegue quindi l'utilizzo di metodi robusti per quantificare l'associazione tra le variabili.

Se i gruppi presenti nelle variabili qualitative sono più di due verrà condotto un test di Kruskal-Wallis per valutarne l'associazione, dal momento che non fornisce informazioni specifiche su quali gruppi differiscano tra loro, si conduce un test post-hoc di confronto multiplo utilizzando la correzione di Bonferroni (Piccolo, 1998).

Si riportano solamente le associazioni risultate significative.

Variabili che comprendono caratteristiche generali del paziente

Si procede esaminando la variabili *eta*, *Eosinofili*, *BmI*, *Dosaggio ICS* per assenza o presenza di SAD, per verificare un'eventuale differenza tra malati e sani (Figura 2.4 e Tabella 2.4).

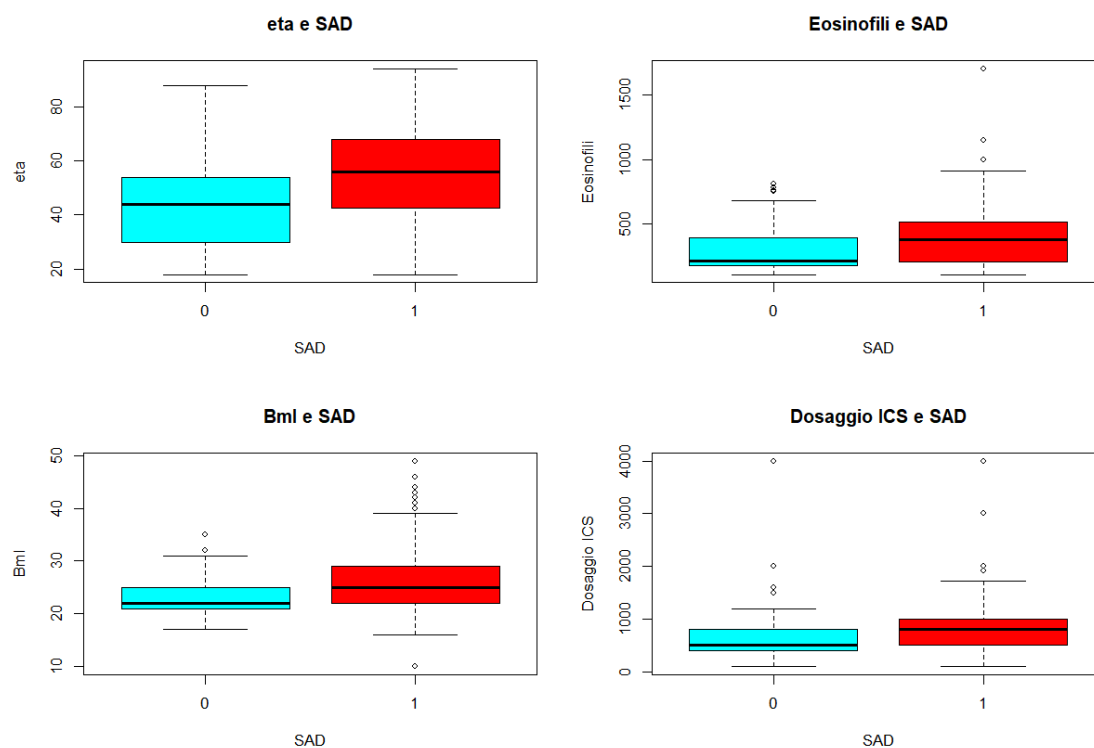


Figura 2.4: Boxplot delle variabili generali del paziente in relazione a SAD.

VARIABILE	SAD = 1		SAD = 0		TEST	P-VALUE
	MEDIANA	IQR	MEDIANA	IQR		
ETA	56.0	25.25	44.0	24	28321	< 0.0001
EOSINOFILI	315.0	161.5	315.0	105	34217	< 0.0001
BMI	25.0	7	22.0	4	26838	< 0.0001
DOSAGGIO ICS	800.0	500	500.0	400	28566	< 0.0001

Tabella 2.4: Mediana e scarto interquantilico delle variabili generali del paziente in relazione a SAD.

Osservando il grafico per la variabile *eta* riportato in Figura 2.4 si nota come chi è affetto da SAD abbia un'età più elevata rispetto a chi non soffre di disfunzioni delle piccole vie aeree. Il test di Mann-Whitney porta al rifiuto dell'ipotesi nulla di uguaglianza delle due distribuzioni ($W = 28321$, $p\text{-value} < 0.0001$). L'età, al momento della diagnosi, non è pertanto omogenea rispetto alle disfunzioni delle piccole vie aeree.

Si nota una presenza in misura maggiore di globuli bianchi coinvolti nelle reazioni allergiche per i pazienti affetti da disfunzione delle piccole vie aeree, il test di Mann-Whitney conferma infatti che le osservazioni nei due gruppi differiscono in modo significativo tra loro ($W = 34217$, $p\text{-value} < 0.0001$).

I pazienti affetti da disfunzioni delle piccole vie aeree sembrano avere un indice di massa corporea maggiore rispetto a quelli non malati di SAD. Il test di Mann-Whitney porta al rifiuto dell'ipotesi nulla di uguaglianza delle due distribuzioni ($W = 26838$, $p\text{-value} < 0.0001$).

Si nota una leggera differenza in variabilità tra le due distribuzioni per il dosaggio di terapia somministrata, a confermare di ciò si conduce il test di Mann-Whitney che porta a rifiutare l'ipotesi di omogeneità tra le due distribuzioni ($W = 28566$, $p\text{-value} < 0.0001$).

Variabili relative alla spirometria

Si procede esaminando la variabile *feno*, *fEV1p*, *fVCp*, *fEV1pfVCp* e *fEf25-75* per assenza o presenza di SAD, per verificare un'eventuale differenza tra malati e sani (Figura 2.5 e Tabella 2.5).

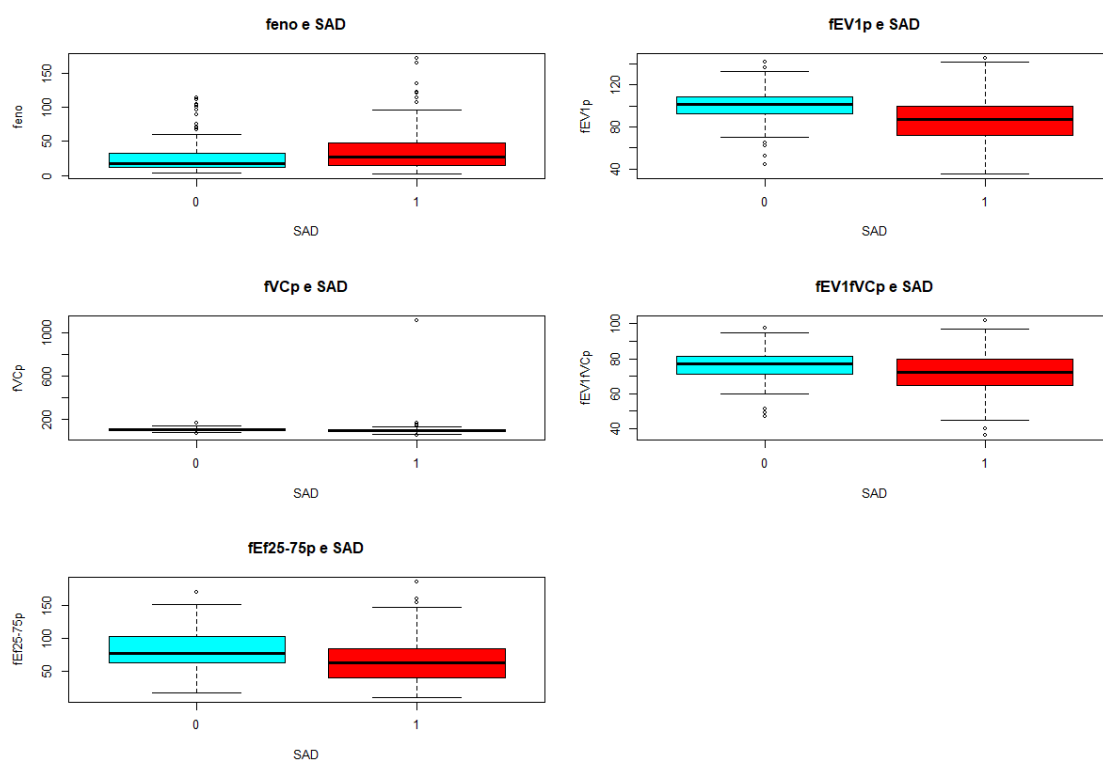


Figura 2.5: Boxplots delle variabili relative ai risultati del test diagnostico spirometrico in relazione a SAD.

VARIABILE	SAD = 1		SAD = 0		TEST	P-VALUE
	MEDIANA	IQR	MEDIANA	IQR		
FENO	26.5	35	20.0	24.8	34056	< 0.0001
FEV1P	87.5	28	101.0	16.5	62115	< 0.0001
FVCP	96.0	20	108.0	17	61499	< 0.0001
FEV1PFVCP	73.0	14.8	76.4	10	53629	< 0.0001
FEF25-75	63.0	44	76.0	39.5	56891	< 0.0001

Tabella 2.5: Mediana e scarto interquantilico delle variabili relative al test spirometrico in relazione a SAD.

Dal boxplot in Figura 2.5 si nota una leggera differenza in variabilità tra le due distribuzioni per la variabile *feno*, infatti chi è affetto da SAD presenta una maggior quantità di monossido di azoto nell'aria espirata. Dato il valore basso del p-value del test di Mann-Whitney ($W = 34056$, $p\text{-value} < 0.0001$) si può concludere che c'è una differenza significativa tra i due gruppi.

Si nota che per i pazienti affetti da disfunzioni delle piccole vie aeree i valori di $fEV1p$ sono più bassi indicando infatti la presenza di un problema ostruttivo. A conferma di quanto appena detto è stato condotto il test di Mann-Whitney che porta ad affermare che vi è una differenza tra i due gruppi presi in considerazione ($W = 62115$, $p\text{-value} < 0.0001$).

I boxplot riportati in Figura 2.5 sembrano evidenziare un comportamento simile tra le due distribuzioni di $fVCp$, il $p\text{-value}$ del test di Mann-Whitney smentisce però questa affermazione ($W = 61835$, $p\text{-value} < 0.0001$).

I due boxplot rappresentati mostrano una differenza in variabilità tra le due distribuzioni per la variabile $fEV1pfVCp$. Il test di Mann-Whitney porta al rifiuto dell'ipotesi nulla di uguaglianza delle due distribuzioni ($W = 53629$, $p\text{-value} < 0.0001$).

I boxplot riportati in Figura 2.5 evidenziano una differenza tra le due distribuzioni di osservazioni di $fEf25-75$. Il test di Mann-Whitney per l'omogeneità delle distribuzioni porta al rifiuto dell'ipotesi nulla ($W = 56891$, $p\text{-value} < 0.0001$).

Variabili relative all'oscillometria forzata

Si procede esaminando la variabili $R5$, $X5$, $R5R20$, AX , $fres$ e $deltaX5$ per assenza o presenza di SAD, per verificare un'eventuale differenza tra malati e sani (Figura 2.6 e Tabella 2.6).

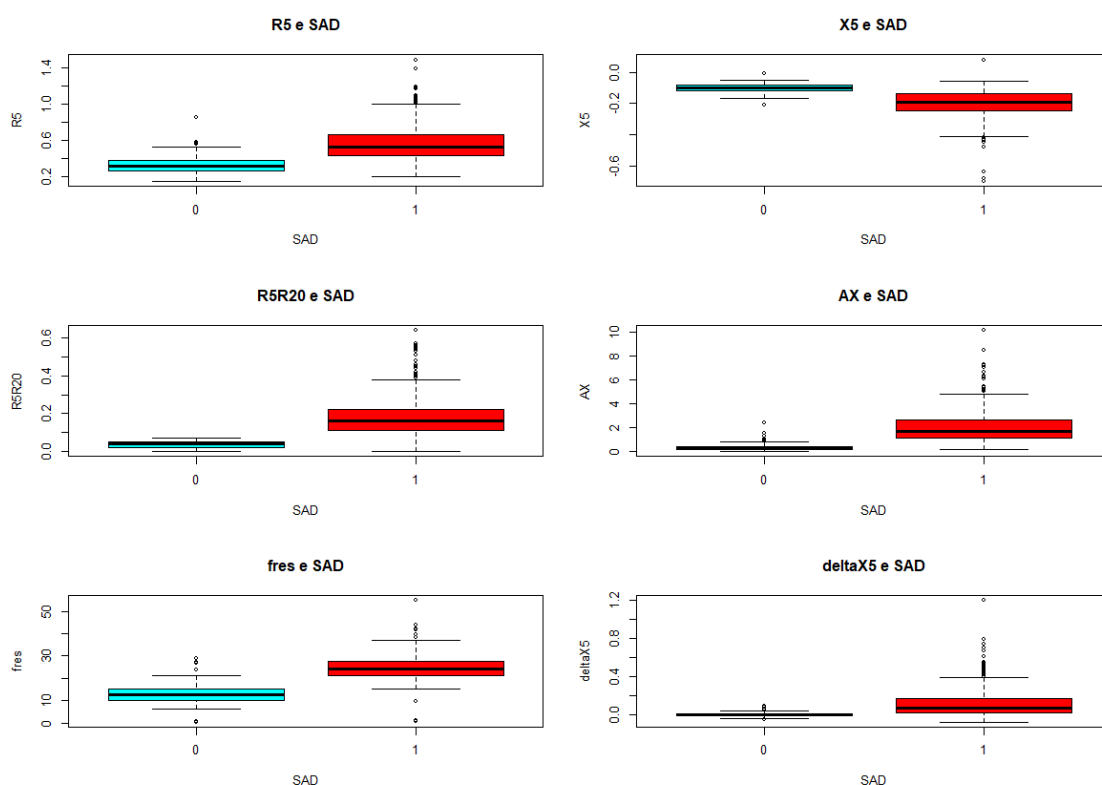


Figura 2.6: Boxplot delle variabili relative ai risultati del test diagnostico basato sull'oscillometria forzata in relazione a SAD.

VARIABILE	SAD = 1		SAD = 0		TEST	P-VALUE
	MEDIANA	IQR	MEDIANA	IQR		
R5	0.53	0.23	0.32	0.11	9140.5	< 0.0001
X5	-0.19	-0.39	-0.10	-0.2	78500	< 0.0001
R5R20	0.16	0.1	0.04	0.03	807	< 0.0001
AX	1.62	1.46	0.34	0.27	1779.5	< 0.0001
FRES	23.62	6.65	12.90	5.1	3317	< 0.0001
DELTAX5	0.06	0.14	0.00	0.02	11063	< 0.0001

Tabella 2.6: Mediana e scarto interquantilico delle variabili relative al test basato sull'oscillometria forzata in relazione a SAD.

Il boxplot della variabile $R5$ (Figura 2.6), secondo la suddivisione dei pazienti disfunzioni delle piccole vie aeree, evidenzia che le due distribuzioni sembrano essere differenti per variabilità. Il test di Mann-Whitney porta al rifiuto dell'ipotesi nulla di uguaglianza delle due distribuzioni ($W = 9140.5$, $p\text{-value} < 0.0001$).

La reattanza a 5 Hz, al momento della diagnosi, è disomogenea rispetto a SAD. Il test di Mann-Whitney porta al rifiuto dell'ipotesi nulla di uguaglianza delle due distribuzioni ($W = 78500$, $p\text{-value} < 0.0001$).

I boxplot riportati in Figura 2.6 per la variabile *R5R20* sono visibilmente diversi in variabilità, a conferma di quanto appena scritto si conduce il test di Mann-Whitney per valutare l'omogeneità delle distribuzioni, si rifiuta infatti l'ipotesi nulla di uguaglianza delle distribuzioni ($W = 807$, $p\text{-value} < 0.0001$).

Per quanto concerne la variabile *AX* il test di Mann-Whitney per l'omogeneità delle distribuzioni porta al rifiuto dell'ipotesi nulla ($W = 1779.5$, $p\text{-value} < 0.0001$).

Il test di Mann-Whitney per la variabile *fres* distinta per SAD porta al rifiuto dell'ipotesi nulla di uguaglianza delle due distribuzioni ($W = 3317$, $p\text{-value} < 0.0001$). La frequenza di risonanza in Hz, al momento della diagnosi, è pertanto disomogenea rispetto a SAD.

Dal grafico riportato in Figura 2.6 si può notare che i pazienti affetti da disfunzioni delle piccole vie aeree presentano una maggior differenza di flusso inspiratorio ed espiratorio rispetto a chi non è affetto da SAD a conferma di ciò il test di Mann-Whitney porta a rifiutare l'ipotesi di omogeneità delle due distribuzioni ($W = 11063$, $p\text{-value} < 0.0001$).

Associazioni significative di alcune variabili con *fumo*

Si procede esaminando la variabili *eta*, *fVCp*, *duratsma*, *BmI* per fumatori e non, per verificare un'eventuale differenza (Figura 2.7 e Tabella 2.7).

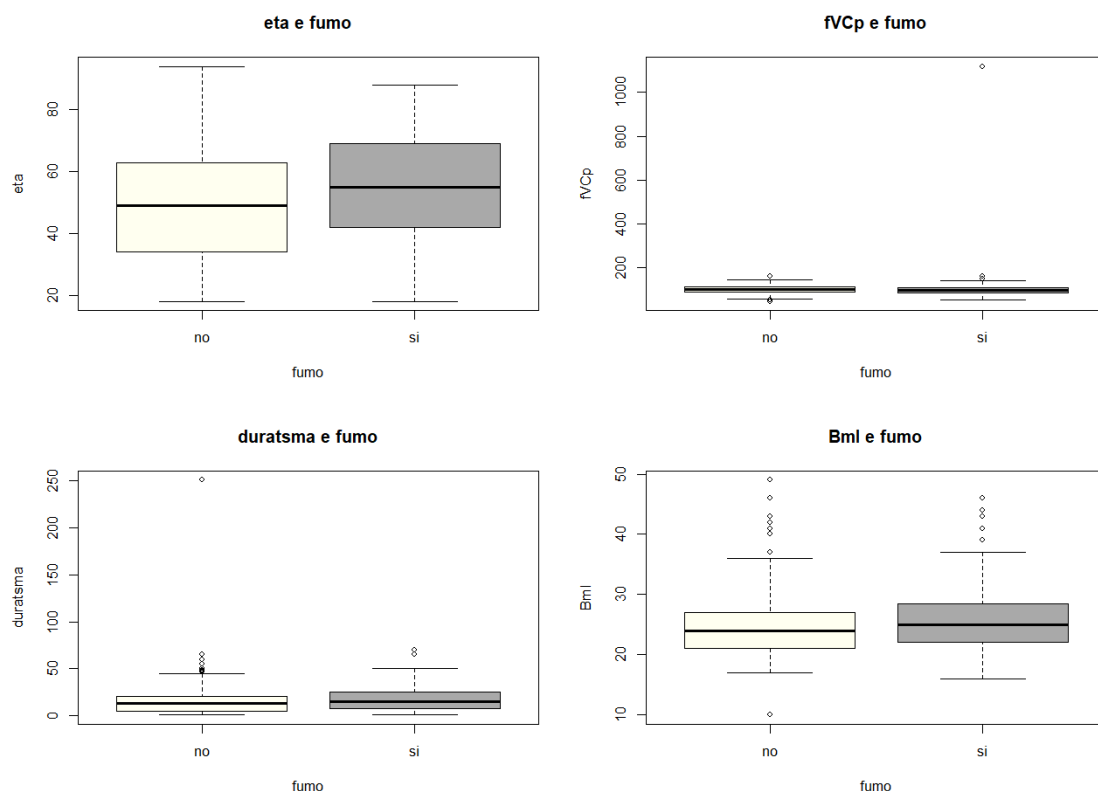


Figura 2.7: Boxplot delle variabili eta, fVCp, duratsma e Bmi relative al tabagismo del paziente.

VARIABILE	FUMO = SI		FUMO = NO		TEST	P-VALUE
	MEDIANA	IQR	MEDIANA	IQR	M-W	
BMI	25.00	6.25	24.00	6	36782	0.006
FVCP	97.5	23.2	101.0	22	47191	0.02634
DURATSMA	15	17.25	13..00	16	38008	0.03032
ETA	55.00	26.5	49.00	29	34374	< 0.0001

Tabella 2.7: Mediana e scarto interquantilico di alcune variabili in relazione al fumo.

L'età per i fumatori risulta essere maggiore, ci sono infatti delle differenze nei due boxplot riportati. Il test di Mann-Whitney per l'omogeneità delle distribuzioni porta al rifiuto dell'ipotesi nulla ($W = 34374$, $p\text{-value} < 0.0001$).

Confrontando il boxplot per presenza o assenza di tabagismo per la variabile $fVCp$, si nota che hanno comportamenti simili. Il test di Mann-Whitney per l'omogeneità delle distri-

buzioni porta però al rifiuto dell'ipotesi nulla ($W = 47191$, $p\text{-value} = 0.02634$).

Per quanto riguarda i boxplot relativi alla variabile *duratsma* si nota che, per fumatori e non fumatori i boxplot sono simili. Il test di Mann-Whitney porta al rifiuto dell'ipotesi nulla ($W = 38008$, $p\text{-value} = 0.03032$).

Si riporta il $p\text{-value}$ del test di Mann-Whitney pari a 0.006 che porta a rifiutare l'omogeneità delle due distribuzioni per la variabile *BmI* distinta per *fumo*.

Relazioni significative tra variabili

Si procede analizzando la variabile *duratsma* con *over65*, la variabile *duratsma* con *AccessiPS* e la variabile *Dosaggio ICS* con *Terapiaextrafine*, per verificare eventuali differenze significative (Figura 2.8). La durata dell'asma è non omogenea rispetto alle due fasce d'età. Il test di Mann-Whitney porta al rifiuto dell'ipotesi nulla di uguaglianza delle due distribuzioni ($W = 25472$, $p\text{-value} < 0.0001$).

Per quanto riguarda i boxplot relativi alla Figura 2.8 si nota che, per chi ha avuto accesso al Pronto Soccorso e chi no i boxplot sono simili. Il test di Mann-Whitney porta al rifiuto dell'ipotesi nulla ($W = 21196$, $p\text{-value} = 0.0111$).

Si vede dalla Figura 2.8 che i pazienti che non sono sottoposti a terapia extrafine necessitano di un Dosaggio Maggiore in mg di terapia. Il test di Mann-Whitney porta al rifiuto dell'ipotesi nulla ($W = 63727$, $p\text{-value} < 0.0001$).

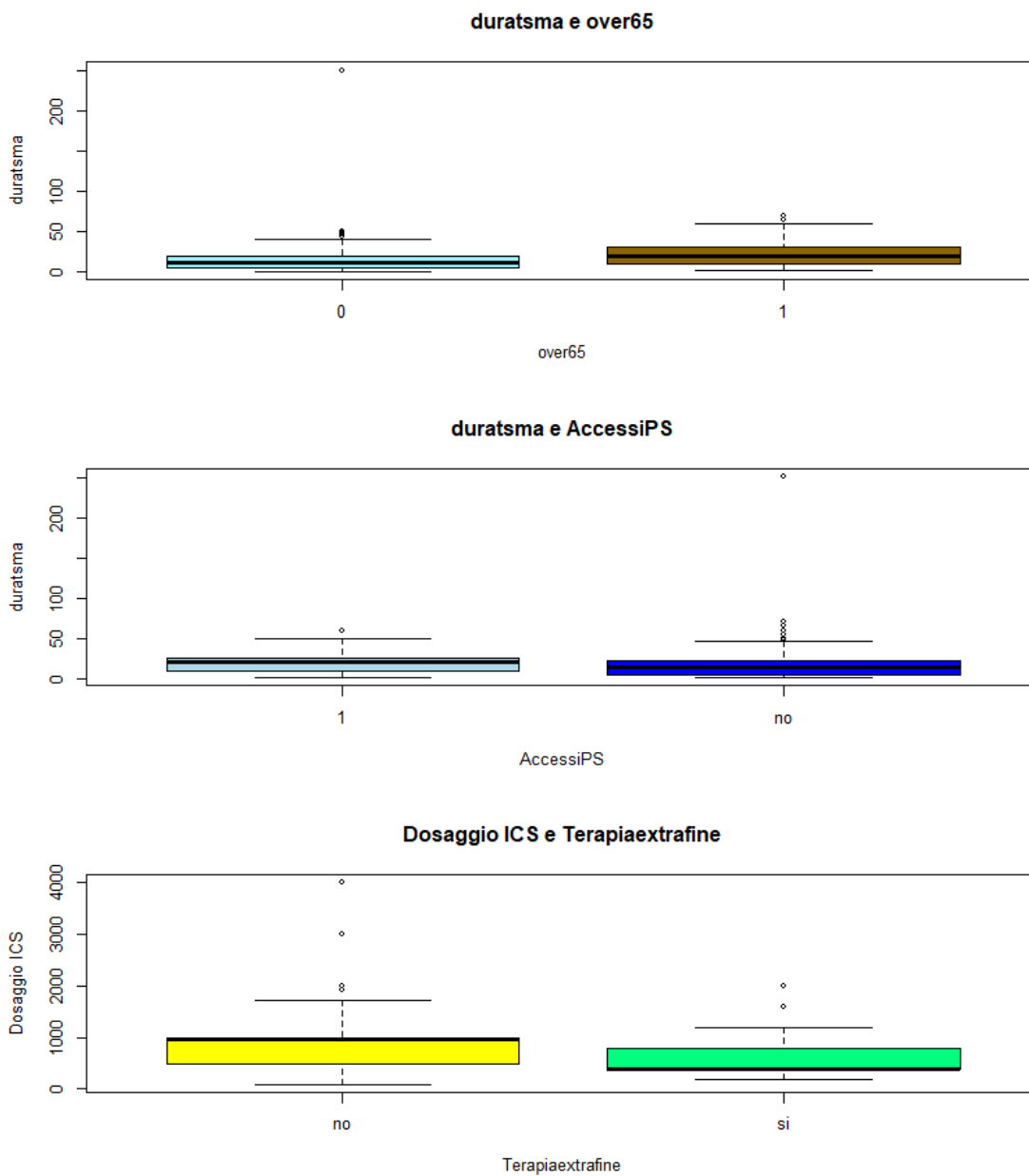


Figura 2.8: Boxplot delle relazioni tra variabili risultate significative.

Associazioni tra variabili con più di due gruppi

Si procede esaminando la variabile *Dosaggio ICS* per i vari livelli di step-terapia e la variabile *Dosaggio ICS* per livelli della variabile *GINA*, in modo da verificare eventuali differenze tra le tre gravità d'asma (Figura 2.9).

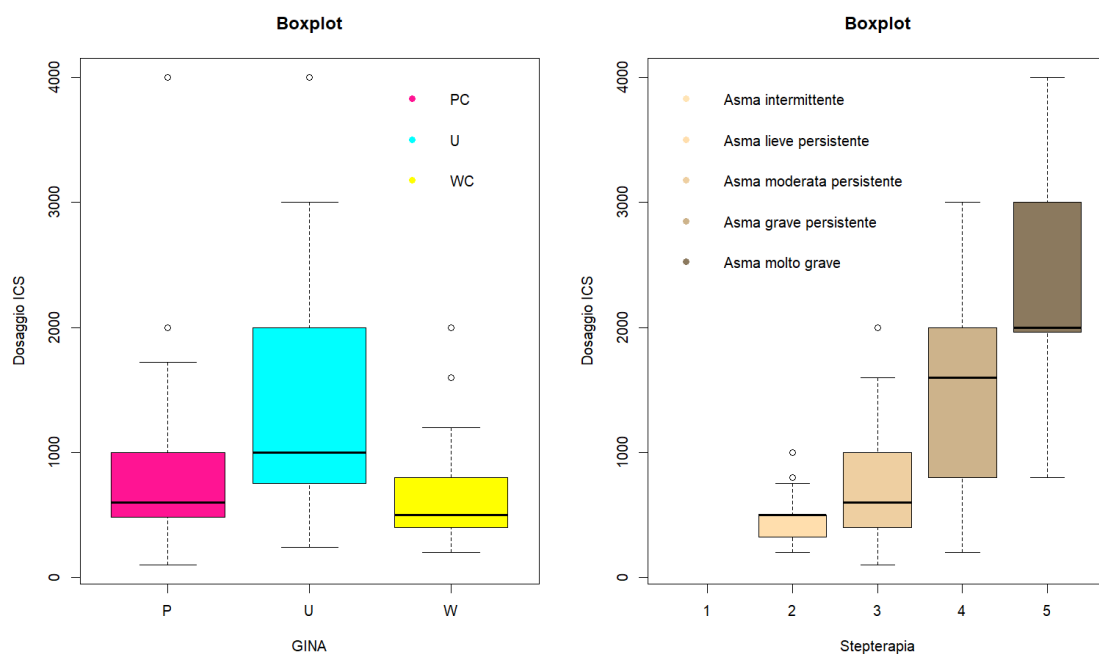


Figura 2.9: Boxplot della variabile Dosaggio ICS distinta per GINA e della variabile Dosaggio ICS distinta per Stepteria.

Dal grafico in Figura 2.9 si vede che man mano che aumenta la gravità dell'asma nei pazienti il dosaggio somministrato è maggiore, a conferma di quanto detto si conduce il test di Kruskal-Wallis che porta ad affermare che vi siano differenze significative tra almeno due gruppi (*Kruskal-Wallis chi-squared* = 101.6231, *df* = 2, *p-value* < 0.0001).

Si conduce ora un'analisi post-hoc con correzione di Bonferroni, ne risulta che tutte le associazioni tra gruppi sono significativamente differenti.

Si procede esaminando la variabile *Dosaggio ICS* per livelli della variabile *GINA*, in modo da verificare eventuali differenze tra le tre gravità d'asma (Figura 2.9). Il boxplot della variabile *Dosaggio ICS* (Figura 2.9), secondo la suddivisione dei pazienti per i livelli di GINA (PC, U e WC), evidenzia che i tre livelli sembrano differenti per variabilità. Il test di Kruskal-Wallis porta al rifiuto dell'ipotesi nulla di ugua-

gianza delle tre distribuzioni (*Kruskal-Wallis chi-squared* = 220.8138, *df* = 3, *p-value* < 0.0001).

Dall'analisi post-hoc con correzione di Bonferroni si può concludere che vi siano differenze statisticamente significative tra tutti i gruppi presenti.

2.2.3 Analisi delle correlazioni

Dal calcolo del coefficiente di correlazione di Spearman tra le varie variabili quantitative (Figura 2.62) emergono alcune correlazioni significative. Tra i più rilevanti, si possono citare quelle positive tra: Eosinofili e feno (0.553), fEf25-75p e fEV1p (0.778), fEf25-75p e fEV1fVCp (0.786) riguardanti la spirometria. Quelle positive tra R5R20 e R5 (0.776), AX e R5 (0.794), AX e R5R20 (0.909), fres e R5R20 (0.763), fres e AX (0.809), deltaX5 e R5R20 (0.731), deltaX5 e AX (0.773) riguardanti l'oscillometria forzata. Si riportano inoltre alcune correlazioni negative risultate significative: X5 e R5 (-0.719), R5R20 e X5 (-0.759), AX e X5 (-0.841), fres e X5 (-0.651), deltaX5 e X5 (-0.679) tutte riguardanti l'oscillometria forzata.

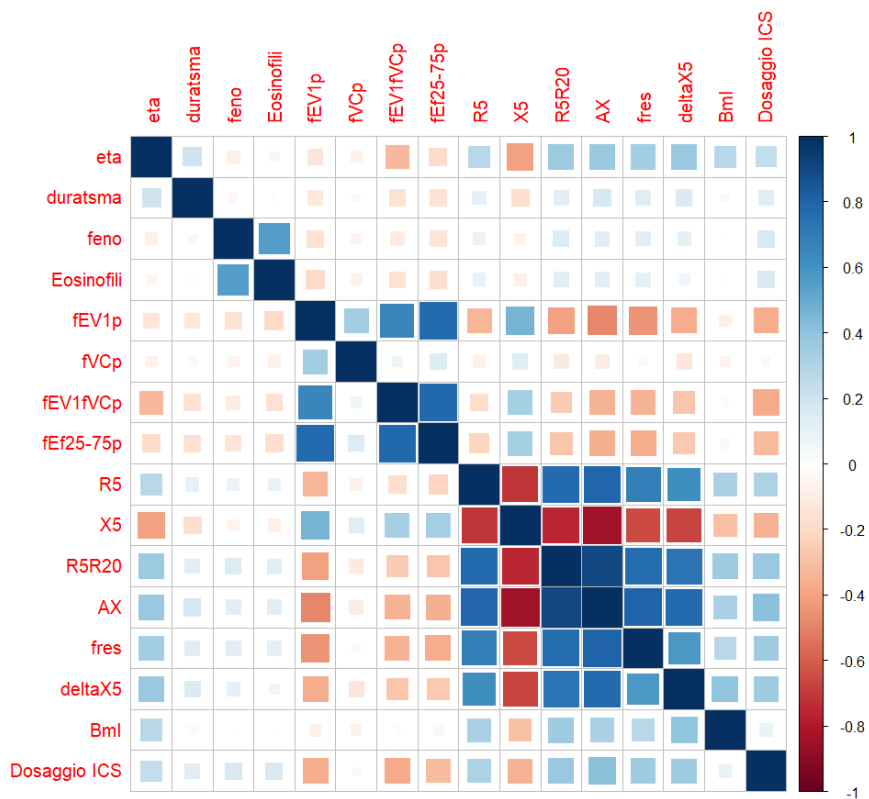


Figura 2.10: Correlazioni di Spearman calcolate per le variabili quantitative.

2.3 Considerazioni finali

Dall'analisi esplorativa delle variabili del *dataset* si sono ottenute diverse informazioni in merito alle varie caratteristiche rilevate sui 618 pazienti che potrebbero essere affetti da disfunzioni delle piccole vie aeree.

Nel prossimo capitolo si effettuerà lo studio sulla variabile *SAD* per esaminare quali fattori possono influenzarla; in particolare si andrà a valutare come questa malattia si comporti tra chi ha meno di sessantacinque anni e chi è invece più anziano.

Capitolo 3

Regressione Logistica

In questo capitolo si indaga circa la presenza di possibili disfunzioni delle piccole vie aeree sui pazienti da parte delle diverse variabili presenti nel *dataset*. Tale ricerca viene effettuata per ogni variabile studiando un modello di regressione logistica che si adatti al meglio ai dati a disposizione.

3.1 Modello di regressione logistica

La regressione logistica appartiene a un'ampia classe di modelli di regressione (GLM). Nei modelli lineari generalizzati vengono attenuate alcune ipotesi fondamentali del modello lineare semplice, quali la linearità della funzione legame tra le variabili, la normalità e l'omoschedasticità delle osservazioni.

La scelta del modello ottimale viene effettuata selezionando le variabili, a partire dal modello con la sola intercetta, con una procedura *forward*, effettuando aggiunte o eliminazioni attraverso il criterio di informazione di Akaike (AIC: *Akaike Information Criterion*) e la significatività delle variabili. La procedura si arresta quando l'aggiunta o l'eliminazione di variabili non migliora l'AIC del modello.

Il calcolo dell'AIC è basato sulla massima verosimiglianza, che misura quanto bene il modello si adatta ai dati osservati. Si utilizza l'AIC poiché con questo criterio si ha un *trade-off* tra adattamento e complessità del modello, per la generalità e la flessibilità.

Il modello di regressione logistica è ampiamente utilizzato in campo medico, biologico, sociale e in molti altri settori per la classificazione e la previsione.

La regressione logistica stima la probabilità del verificarsi di un evento sulla base di uno specifico *dataset* di variabili indipendenti (o predittori). Poiché il risultato è una probabilità, la variabile dipendente è vincolata tra 0 e 1. Nella regressione logistica, viene applicata una trasformazione *logit* sulle probabilità (la probabilità di successo divisa per la probabilità di fallimento).

- In un modello di regressione logistica la variabile dipendente Y è una variabile Bernoulliana:

$$P(Y_i = x) = \begin{cases} \pi_i & \text{se } x = 1 \\ 1 - \pi_i & \text{se } x = 0 \end{cases} \text{ con media } E(Y_i) = \pi_i$$

- Si ha $g(\pi_i) = \beta_1 x_{i1} + \dots + \beta_p x_{ip}$ con g funzione legame.
- La funzione legame può essere di tipo *logit* oppure *probit*, nel caso di regressione logistica si usa la funzione *logit*:
 $\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right)$

con

$$\pi_i = \frac{e^{\beta^T x_i}}{1 + e^{\beta^T x_i}}$$

- Si ha eteroschedasticità:

$$\text{Var}(Y_i) = \pi_i(1 - \pi_i), \quad i = 1, \dots, n$$

- Il modello di regressione logistica prevede numerosi vantaggi come l'interpretabilità, semplificazioni nelle quantità di verosimiglianza e la possibilità di trattare dati raccolti retrospettivamente (Agresti, 2018).

Per la verifica della bontà di adattamento del modello selezionato con i dati a disposizione viene creata la tabella di corretta classificazione utilizzata per valutare le prestazioni di un modello di classificazione. Si costruisce il grafico della curva ROC e si calcola l'AUC (*Area Under the Curve*) per misurare la bontà complessiva della curva ROC. Si conduce infine il test di Hosmer-Lemeshow che verifica se le stime del modello sono simili alle osservazioni effettive.

Di seguito vengono riportati i modelli stimati, accompagnati dal valore della Devianza Residua, dai gradi di libertà e dal livello di significatività del test di Hosmer-Lemeshow sulla bontà di adattamento del modello.

3.2 Modellazione

L'analisi di regressione logistica comincia dunque con il modello nullo composto dalla sola intercetta.

Inizia l'iterazione per selezionare le variabili una alla volta in base alla significatività delle variabili e al fatto che diminuiscano o meno l'AIC del modello corrente: AX (p-value < 0.0001 e $AIC = 236.92$) risulta la variabile di partenza per iniziare il processo *forward* di inserimento delle variabili .

L'inserimento della variabile AX porta ad una diminuzione dell'AIC del modello che risulta ora pari a 236.92 con una Devianza Residua di 232.92 e 616 gradi di libertà.

Si valuta nuovamente se l'introduzione nel modello di regressione logistica di altre variabili possa migliorare ulterior-

mente l'adattamento del modello ai dati. La variabile *EIA* risulta essere significativa (p-value < 0.0001 e AIC = 211.87).

Si valuta l'aggiunta di altre variabili al modello. Risulta essere significativa la variabile *deltaX5* con un p-value pari a 0.0025.

Risulta significativa anche la variabile *over65* con un p-value pari a 0.0235.

Infine l'ultima variabile a risultare significativa è *Riacutizzazioni* con un p-value pari a 0.02.

Si compone così il modello finale con un AIC pari a 198.27 e una Devianza Residua uguale a 186.27 con 612 gradi di libertà.

3.2.1 Modello finale

Dopo aver valutato una a una le variabili significative da aggiungere al modello nullo si arriva al modello completo finale contenente le variabili *AX*, *EIA*, *deltaX5*, *over65* e *Riacutizzazioni*, rappresentato in Tabella 3.1.

VARIABILE	Estimate	Std. Error.	z value	Pr(> Z)
INTERCETTA	-5.434	0.548	-9.908	< 0.0001
AX	6.948	0.796	8.732	< 0.0001
EIA - si	1.667	0.437	3.814	0.0002
DELTA5	17.560	6.346	2.767	0.006
OVER65 - si	-1.458	0.585	-2.494	0.012
RIACUTIZZAZIONI - si	1.146	0.497	2.303	0.021

Tabella 3.1: Modello di regressione logistica finale.

Il modello finale presenta tutti i coefficienti significativi a un livello di α fissato pari a 0.05, ha Devianza Residua pari a 186.27 con 612 gradi di libertà e un AIC di 198.27.

3.2.2 Interpretazione dei coefficienti

I coefficienti di un modello di regressione logistica indicano il contributo relativo delle variabili indipendenti nella previsione della variabile risposta *SAD*.

- Il coefficiente per la variabile quantitativa *AX* è pari a 6.948, aumentare il valore dell'area di reattanza di una unità vorrebbe quindi dire aumentare il log-odds di avere disfunzioni delle piccole vie aeree di quasi 7 unità quando tutte le altre variabili sono costanti.
- Il coefficiente per *EIA* è di 1.667: se un individuo presenta asma da sforzo il log-odds di avere disfunzioni delle piccole vie aeree aumenta di 1.7 unità rispetto a un individuo che non ha asma da sforzo, mantenendo costanti le altre variabili del modello.
- Il coefficiente per *deltaX5* è 17.56. Aumentare il valore della differenza tra flusso inspiratorio e espiratorio a 5 Hz di una unità aumenta il log-odds di avere disfunzioni delle piccole vie aeree di 18 unità quando tutte le altre variabili sono costanti.
- Il coefficiente per *over65* è negativo (-1.458), questo valore indica che se un individuo ha più di 65 anni (*over65* = 1), il log-odds di successo (*SAD* = 1) diminuisce di 1.4581 unità rispetto a un individuo che ha meno di 65 anni (*over65* = 0), mantenendo costanti tutte le altre variabili nel modello.
- Il coefficiente per la variabile *Riacutizzazioni* è pari a 1.146, ciò significa che presentare riprese dell'andamento

acuto dell'asma da parte di un paziente indica che il log-odds di avere disfunzioni delle piccole vie aeree aumenta di circa 1 unità rispetto a chi non presenta riacutizzazioni, quando tutte le altre variabili sono costanti.

3.3 Bontà di adattamento del modello

Il valore dell'AIC del modello finale pari a 198.27 è molto minore rispetto al modello nullo (AIC = 799.36) e indica quindi che il modello finale sia migliore rispetto al modello nullo e a tutti i modelli correnti considerati aggiungendo volta per volta le variabili.

Si conduce dunque il Test del Rapporto di Verosimiglianza per confrontare modello nullo e modello finale, il valore del p-value molto vicino allo zero (p-value < 0.0001) indica che il modello finale è significativamente migliore del modello nullo. Questo suggerisce che le variabili indipendenti elencate nel Modello 2 hanno un effetto significativo sul modello di regressione logistica e migliorano la sua capacità di adattamento ai dati.

Si costruisce ora la tabella di corretta classificazione per valutare le prestazioni del modello, questa tabella fornisce informazioni sulle previsioni del modello rispetto ai valori osservati dell'insieme di dati preso in considerazione.

	FALSE	TRUE
SAD = no	201	13
SAD = si	24	380

Tabella 3.2: Tabella di corretta classificazione per il modello finale di regressione logistica.

Vi sono dunque 201 casi in cui il modello ha previsto correttamente l'assenza di disfunzioni delle piccole vie aeree e

380 casi in cui il modello ha previsto correttamente la presenza di *SAD*. Vi sono invece 13 falsi positivi e 24 falsi negativi.

Si utilizzano queste informazioni per calcolare alcune metriche di valutazione delle prestazioni del modello, come l'accuratezza, che misura la frazione di previsioni corrette rispetto al totale delle previsioni, dunque l'accuratezza del modello è circa del 94%:

$$\frac{TP + TN}{TP + TN + FP + FN} = \frac{380 + 201}{380 + 201 + 13 + 24} \approx 0.94.$$

Si costruisce ora la curva ROC e si calcola l'AUC per valutare ulteriormente le prestazioni del modello costruito (Figura 3.2).

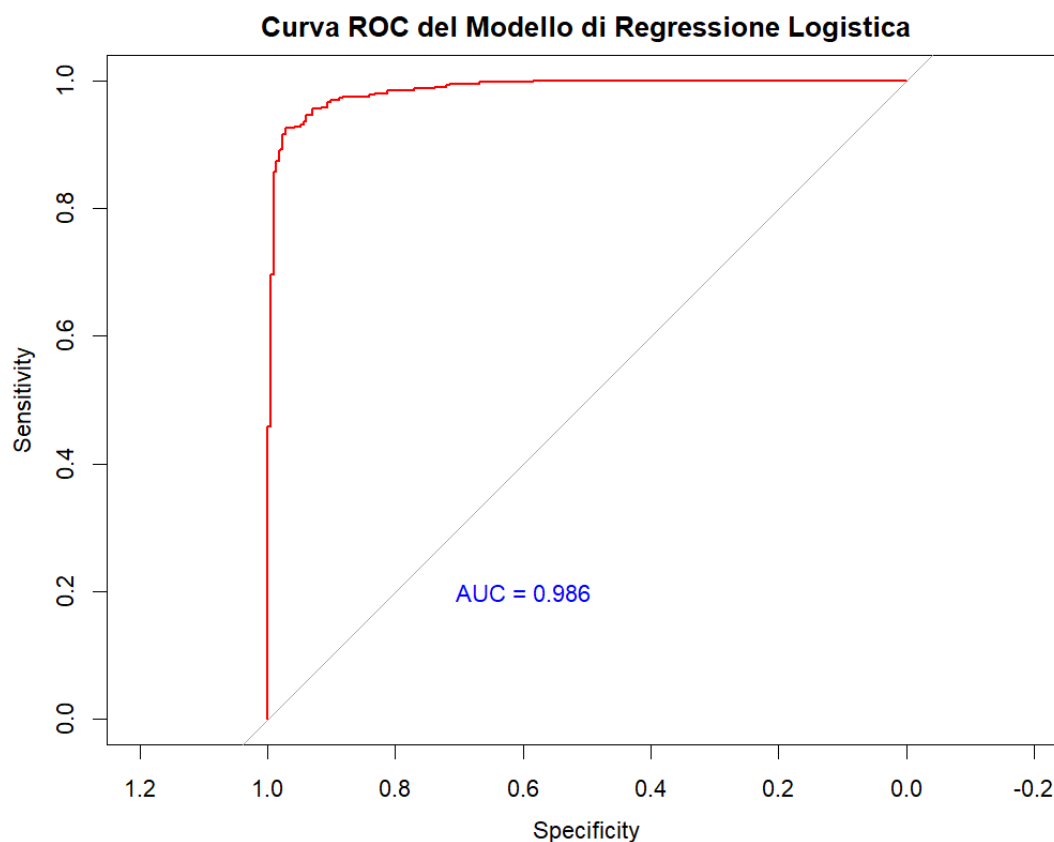


Figura 3.1: Curva ROC e AUC del modello di regressione logistica finale.

La curva ROC mostra il tasso di veri positivi sulle ordinate rispetto al tasso di falsi positivi sulle ascisse al variare delle soglie di classificazione. La curva ROC rappresentata in Figura 3.2 si avvicina molto all'angolo in alto a sinistra indice che il modello ha una buona sensibilità con pochi falsi positivi.

L'AUC rappresenta l'area sottesa dalla curva ROC e misura la capacità complessiva del modello di discriminare tra le classi, il valore dell'AUC del modello preso in considerazione è pari a 0.986.

Infine per confermare la bontà di adattamento del modello ai dati si conduce il test di Hosmer-Lemeshow che rappresenta uno strumento per la valutazione della bontà di adattamento di un modello basato sul test del chi-quadro. Nel test di H-L l'ipotesi nulla si basa sul fatto che il modello di regressione logistica valuti in modo accurato le probabilità rispetto alle osservazioni effettive. Il valore del p-value porta a respingere l'ipotesi nulla, c'è quindi evidenza statistica che il modello finale non si adatti bene ai dati osservati ($X^2 = 618$, $df = 8$, $p\text{-value} < 0.0001$).

Il risultato dato dal test di H-L si discosta dalle conclusioni ottenute dalla curva ROC e dal valore dell'AUC, infatti il modello potrebbe essere buono nella discriminazione tra le classi (come misurato dalla curva ROC e dall'AUC) ma potrebbe avere problemi di calibrazione, cioè le probabilità previste potrebbero non riflettere accuratamente le probabilità osservate. In questo caso, si ha una buona discriminazione ma una cattiva calibrazione come dimostrato dal test di Hosmer-Lemeshow.

3.4 Considerazioni finali

Dal modello di regressione logistica stimato in questo capitolo le variabili risultate significative sono *AX*, *EIA*, *deltaX5*, *over65* e *Riacutizzazioni*, è d'interesse notare come solo alcuni parametri del test basato sull'oscillometria forzata siano risultati significativi, mentre nessun parametro spirometrico risulta entrare a far parte del modello.

Inoltre *Terapiaextrafine* e *Dosaggio ICS* che indicano rispettivamente il tipo di terapia somministrato al paziente e il dosaggio di quest'ultima in mg non entrano mai a far parte del modello, indice che probabilmente non vi è alcuna differenza tra il tipo di terapia al quale il paziente è sottoposto e alla quantità di mg somministrati.

La curva ROC e il valore dell'AUC indicano che il modello si adatta bene ai dati, ma il test di Hosmer-Lemeshow smen-tisce la bontà di adattamento del modello, questo potrebbe essere dovuto a problemi di calibrazione del modello. Il modello stimato sembra quindi discriminare bene i dati ma non sembra essere ben calibrato.

Nel prossimo capitolo verrà utilizzata la tecnica statistica del *Propensity Score Matching*, utile per ridurre il *confounding* e rendere più robusti i risultati dello studio.

Capitolo 4

Propensity Score Matching

Il *Propensity Score Matching* è una tecnica statistica utilizzata nell'analisi dei dati per ridurre i bias da selezione negli studi clinici e serve a rendere dati osservazionali “assimilabili” a dati randomizzati. Il *Propensity Score* è una stima della probabilità che un paziente sia assegnato a un gruppo piuttosto che all'altro, sulla base di una serie di covariate che potrebbero influenzare la scelta del gruppo. In altre parole, il *Propensity Score* rappresenta la "propensione" di un individuo a far parte di un gruppo in base alle caratteristiche che lo rappresentano (Scardapane, 2023).

Una volta calcolati i *Propensity Scores* per tutti i pazienti dello studio, è possibile utilizzare questa tecnica per creare un campione di controllo che sia comparabile al gruppo di trattamento. Questo viene fatto abbinando ciascun paziente del gruppo di trattamento a un individuo del gruppo di controllo con *Propensity Scores* simili.

È una tecnica utile per ridurre il confounding e rendere più robusti i risultati degli studi osservazionali quando non è possibile condurre studi sperimentali randomizzati.

4.1 Applicazione al dataset

Si procede calcolando i *Propensity Scores* per la variabile *over65* che indica se il paziente preso in considerazione ha più o meno di sessantacinque anni condizionatamente alle variabili *fumo*, *BmI* e *duratsma*.

Prima dell'applicazione del *matching* vi sono 456 unità nel gruppo di controllo e 162 unità nel gruppo trattato. Dopo l'applicazione di questa tecnica si hanno 162 unità nel gruppo di controllo (più giovani di 65 anni) e 162 unità nel gruppo trattato (più anziani di 65 anni), ci sono poi 294 unità non abbinate nel gruppo di controllo.

In Figura 4.1 viene rappresentato il grafico *jitter* dove ogni punto del grafico corrisponde a un paziente.

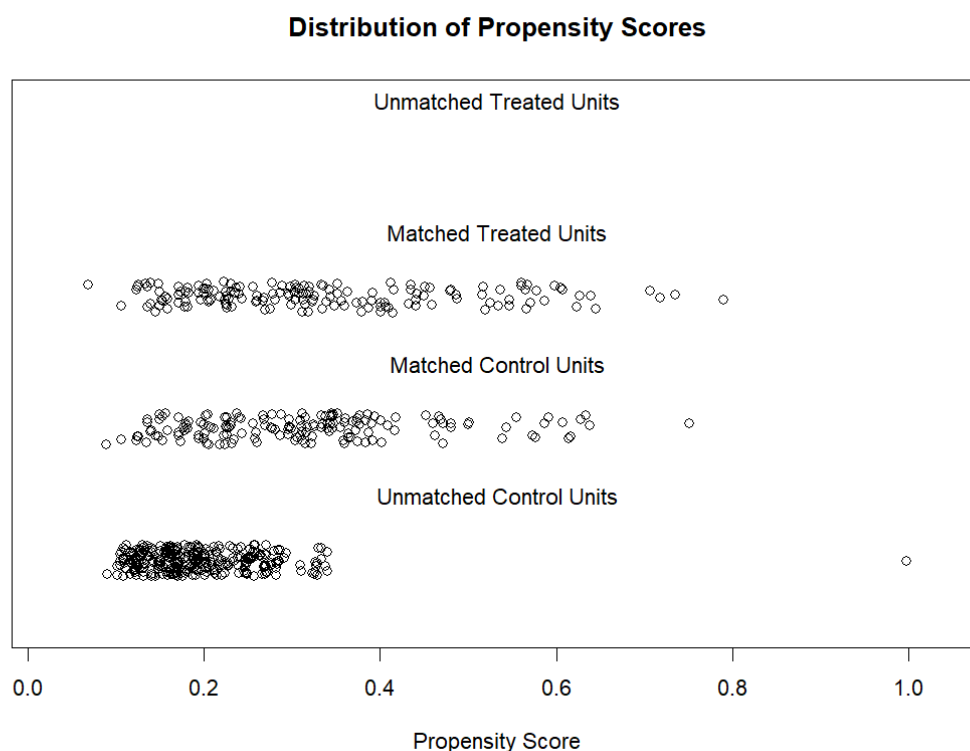


Figura 4.1: Rappresentazione dei dati dopo il *Propensity Score Matching*.

Si confrontano quindi gli indicatori principali delle covariate tra il gruppo di trattamento ($over65 = 1$) e il gruppo di controllo ($over65 = 0$) prima del *matching* (Tabella 4.1) con quelli delle covariate dopo il *matching* (Tabella 4.2).

VARIABILE	Means Treated	Means Control	Std. Mean Diff.	Var. Ratio	eCDF Mean	eCDF Max
DISTANCE	0.3325	0.2371	0.6258	1.8088	0.2067	0.3245
FUMO - no	0.5617	0.7018	-0.2822	.	0.1400	0.1400
FUMO - si	0.4383	0.2982	0.2822	.	0.1400	0.1400
BMI	26.7407	24.5636	0.3794	1.4075	0.0713	0.2223
DURATSMA	22.6790	15.1096	0.4733	1.0207	0.1382	0.2561

Tabella 4.1: Statistiche di bilanciamento delle variabili tra il gruppo trattato e il gruppo di controllo prima dell'applicazione del *matching*.

Nella Tabella 4.1 vengono riportate le medie delle covariate per il gruppo trattato e di controllo (Means Treated e Means Control), la differenza media standardizzata tra i gruppi trattati e di controllo (Std. Mean Diff.), il rapporto tra le varianze delle covariate (Var. Ratio) e la differenza media e massima delle funzioni di distribuzione cumulativa empirica (eCDF Mean e eCDF Max). Nella Tabella 4.2 viene inoltre riportata la distanza standard tra coppie (Std. Pair Dist.)

Questa sezione presenta le stesse statistiche, ma solamente per i dati che sono stati effettivamente abbinati.

VARIABILE	Means Treated	Means Control	Std. Mean Diff.	Var. Ratio	eCDF Mean	eCDF Max	Std. Pair Dist.
DISTANCE	0.3325	0.3128	0.1290	1.3684	0.0166	0.1235	0.1321
FUMO - no	0.5617	0.5802	-0.0373	.	0.0185	0.0185	0.7589
FUMO - si	0.4383	0.4198	0.0373	.	0.0185	0.0185	0.7589
BMI	26.7407	26.6173	0.0215	0.9755	0.0211	0.0864	0.7186
DURATSMA	22.6790	20.6914	0.1243	1.4825	0.0359	0.1049	0.6654

Tabella 4.2: Statistiche di bilanciamento delle variabili tra il gruppo trattato e il gruppo di controllo dopo dell'applicazione del *matching*.

Confrontando le due tabelle si nota che le statistiche sono più bilanciate in Tabella 4.2 rispetto a quelle prima del *matching*. Osservando lo *Standardized Mean Difference* si vede che i valori riportati in Tabella 4.2 sono molto più vi-

cini al valore 0 rispetto ai valori dello stesso prima del *matching*. Questo sta a indicare che le due popolazioni (gruppo di trattamento e di controllo) sono più simili tra loro dopo il *matching* per la variabile considerata. Si osservano inoltre i *Variance Ratio*, cioè il rapporto delle varianze tra il gruppo trattato e il gruppo di controllo per una specifica variabile. Un valore vicino a 1 indica che le varianze nei due gruppi sono simili. Nelle covariate dopo il *matching* tutti i valori del *Variance Ratio* si avvicinano di più al valore unitario, tranne per la variabile *duratsma*, dove il *Variance Ratio* sembra discostarsi maggiormente dopo il *matching*.

In generale, il *matching* sembra aver contribuito a migliorare il bilanciamento tra i gruppi *over65* e di controllo, ma alcune differenze rimangono per la variabile *duratsma*.

Si visualizza quindi il bilanciamento ottenuto con il *Propensity Score Matching* con il grafico di densità per le variabili *fumo*, *BmI* e *duratsma* (Figura 4.2).

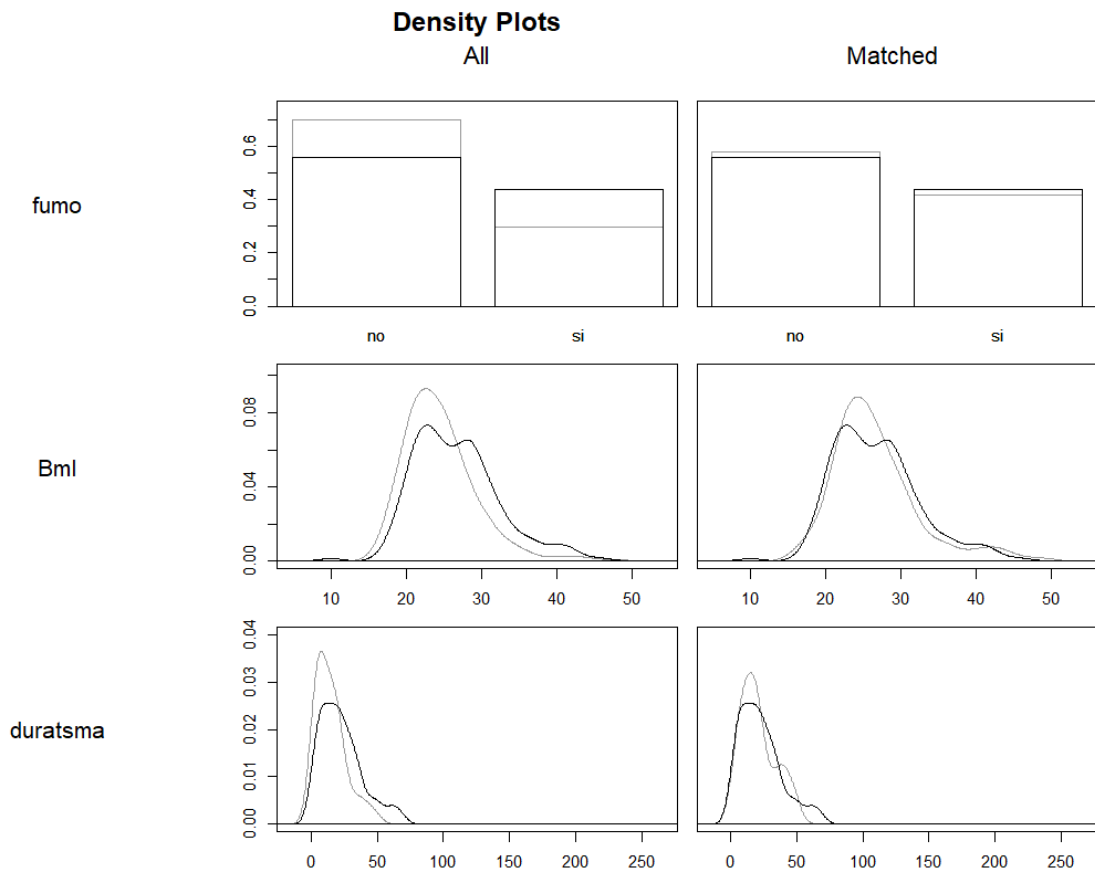


Figura 4.2: Grafico delle densità delle variabili fumo, BmI e duratsma dopo l'applicazione del *Propensity Score Matching*.

Il grafico riportato in Figura 4.2 rappresenta la distribuzione delle variabili selezionate per entrambi i gruppi: trattamento, cioè chi ha più di 65 anni (in nero) e controllo (in grigio). L'obiettivo di questo grafico è valutare visivamente se le distribuzioni delle variabili selezionate sono simili tra il gruppo di trattamento e il gruppo di controllo dopo il *PSM*.

Per le variabili *fumo* e *BmI* le curve di densità dei due gruppi (nero e grigio) sono molto simili e si sovrappongono, ciò indica un buon bilanciamento tra i gruppi per la variabili considerate. Per la variabile *duratsma* invece anche dopo il *matching* le curve di densità mostrano una sovrapposizione incompleta.

4.2 Analisi dei risultati

In questa sezione ci si propone di analizzare nuovamente i dati a disposizione dopo aver effettuato il *Propensity Score Matching* in modo da valutare eventuali cambiamenti dall'analisi effettuata senza appaiamento.

4.2.1 Analisi bivariata

Di seguito si riporta l'analisi bivariata dei dati in relazione a *over65* per valutare quali relazioni risultino significative e in che modo ogni paziente si comporti in relazione all'età (più di 65 anni o meno di 65 anni).

Variabili qualitative distinte per *over65*

In Tabella 4.3 vengono riportate le frequenze assolute e percentuali delle variabili qualitative distinte per i livelli della variabile *over65*. Viene inoltre condotto il test chi-quadro per valutare il grado di associazione tra le varie variabili e *over65*.

VARIABILE	CATEGORIE	OVER65 = 0	OVER65 = 1	P-VALUE
SAD	Si	110 (67.9%)	133 (82.1%)	0.0047
	No	52 (32.1%)	29 (17%)	
SESSO	Femmine	83 (51.2%)	97 (59.8%)	0.146
	Maschi	79 (48.8%)	65 (40.2%)	
ATOPIA	Si	108 (66.7%)	67 (41.3%)	< 0.0001
	No	54 (33.3%)	95 (58.64%)	
GINA	WC	46 (28.4%)	39 (24.1%)	0.617
	PC	72 (44.4%)	73 (45.1%)	
	U	44 (27.2%)	50 (30.9%)	
GINA1	WC	46 (28.4%)	39 (24.1%)	0.449
	PC-U	116 (71.6%)	123 (75.9%)	
RIACUTIZZAZIONI	Si	64 (39.5%)	84 (51.9%)	0.03
	No	98 (60.5%)	78 (48.1%)	
ACCESSIPS	Si	17 (10.5%)	22 (13.6%)	0.4946
	No	145 (89.5%)	140 (86.4%)	
RICOVERI	Si	5 (3.1%)	8 (4.9%)	0.571
	No	157 (96.9%)	154 (95.1%)	
STEPTERAPIA	Intermittente	5 (3.1%)	4 (2.5%)	<0.0001
	Lieve	37 (22.8%)	8 (4.9%)	
	Moderato	78 (48.1%)	87 (53.7%)	
	Grave	40 (24.7%)	56 (34.6%)	
	Molto grave	2 (1.2%)	7 (4.3%)	
TERAPIAEXTRAFINE	Si	58 (35.8%)	48 (29.6%)	0.2866
	No	104 (64.2%)	114 (70.4%)	
FUMO	Si	68 (41.9%)	71 (43.8%)	0.8224
	No	94 (58.1%)	91 (56.2%)	
RISVEGLI	Si	75 (46.3%)	89 (54.9%)	0.149
	No	87 (53.7%)	73 (45.1%)	
EIA	Si	94 (58.0%)	101 (62.3%)	0.4959
	No	68 (42%)	61 (37.7%)	

Tabella 4.3: Frequenze assolute, frequenze relative e p-value dato dal test del chi-quadro relative delle variabili qualitative distinte per over65 dopo il *Propensity Score Matching*.

Dai risultati ottenuti si può notare che le associazioni che non risultano significative con la variabile *over65* sono quelle relative a *sezzo*, *GINA*, *GINA1*, *AccessiPS*, *Ricoveri*, *Terapiaextrafine*, *fumo*, *Risvegli* e *EIA*, con un p-value maggiore di 0.05. C'è quindi evidenza statistica per concludere che l'età *over65* non sia associata in modo significativo a queste variabili.

Variabili quantitative distinte per *over65*

Si procede esaminando la variabili *duratsma*, *Eosinofili*, *BmI* e *Dosaggio ICS* per chi ha più o meno di 65 anni, per verificare un'eventuale differenza tra giovani e anziani (Tabella 4.4).

VARIABILE	OVER65 = 0		OVER65 = 1		TEST	P-VALUE
	MEDIANA	IQR	MEDIANA	IQR	M-W	
DURATSMA	18.5	19.5	20.0	21	12518	0.4732
EOSINOFILI	315.0	167.5	315.0	137.8	14925	0.029
BMI	25.5	6	26.0	7	12798	0.701
DOSAGGIO ICS	600.0	600	960	900	9462.5	< 0.0001

Tabella 4.4: Mediana e scarto interquantilico delle variabili generali del paziente in relazione a *over65* dopo il *Propensity Score Matching*.

Dal test di Mann-Withney condotto su queste variabili le associazioni a non risultare significative con la variabile *over65* sono quelle con *duratsma* e *BmI*.

Anche prima di applicare il *Propensity Score Matching* l'associazione tra *over65* e *BmI* non risultava significativa, mentre la relazione tra *over65* e *duratsma* si. Le associazioni di *over65* con *Eosinofili* e *Dosaggio ICS* prima dell'appaiamento non risultavano significative.

Variabili relative alla spirometria distinte per *over65*

Si procede esaminando la variabili *feno*, *fEV1p*, *fVCp*, *fEV1pfVCp*, *fEf25-75* per chi ha più o meno di 65 anni, per verificare un'eventuale differenza tra giovani e anziani (Tabella 4.5).

VARIABILE	OVER65 = 0		OVER65 = 1		TEST	P-VALUE
	MEDIANA	IQR	MEDIANA	IQR	M-W	
FENO	24.0	33.75	20.0	24.75	15372	0.0076
FEV1P	90.0	24.75	87.0	32.75	13713	0.4836
FVCP	99.0	22	98.5	28	13420	0.7241
FEV1PFVCP	74.0	11.8	68.9	14.49	16046	0.0005
FEF25-75	66.0	39	54.5	43	15154	0.015

Tabella 4.5: Mediana e scarto interquantilico delle variabili relative al test spirometrico in relazione a *over65* dopo il *Propensity Score Matching*.

Le associazioni con *over65* a non risultare significative sono quelle con *fEV1p* e *fVCp* dato il p-value maggiore di 0.05, così come prima di applicare l'appaiamento ai dati.

Prima dell'applicazione del *Propensity Score Matching* nessuna associazione di *over65* con i risultati del test spirometrico è risultata significativa.

Variabili relative all'oscillometria forzata distinte per *over65*

Si procede esaminando la variabili *R5*, *X5*, *R5R20*, *AX*, *fres* e *deltaX5* per chi ha più o meno di 65 anni, per verificare un'eventuale differenza tra giovani e anziani (Tabella 4.6).

VARIABILE	OVER65 = 0		OVER65 = 1		TEST	P-VALUE
	MEDIANA	IQR	MEDIANA	IQR	M-W	
R5	0.48	0.21	0.53	0.23	10900	0.008
X5	-0.15	-0.33	-0.21	-0.43	17299	< 0.0001
R5R20	0.115	0.14	0.165	0.13	9780	< 0.0001
AX	1.16	1.67	1.96	1.8	9163	< 0.0001
FRES	21.32	10.67	24.37	7.94	9955	0.0002
DELTAX5	0.03	0.11	0.12	0.21	9918.5	0.0001

Tabella 4.6: Mediana e scarto interquantilico delle variabili relative al test basato sull'oscillometria forzata in relazione a *over65* dopo il *Propensity Score Matching*.

Tutte le variabili relative al test dell'oscillometria forzata risultano essere significative in relazione alla variabile *over65*,

c'è quindi evidenza statistica che l'età (*over65*) sia associata in modo significativo ad ogni risultato del test oscillometrico.

Prima dell'applicazione del *Propensity Score Matching* ai dati queste associazioni non risultavano significative.

4.2.2 Modello di regressione logistica con il PSM

Si stima ora il modello di regressione logistica ripetendo la procedura *forward* di selezione delle variabili, si parte quindi dal modello nullo e via via si aggiungono le variabili risultate significative, ci si ferma quando l'AIC del modello non migliora ulteriormente.

L'AIC del modello nullo è pari a 366.4, si parte aggiungendo al modello la variabile *AX* (p-value < 0.0001) che porta l'AIC a diminuire fino a 118.4, aggiungendo poi la variabile *deltaX5* (p-value = 0.0008) si arriva a un AIC pari a 109.2 del modello corrente. Si inserisce la variabile relativa all'asma da sforzo (p-value = 0.0066) che porta l'AIC del modello corrente a 103.8. Si aggiunge infine la variabile *BmI* (p-value = 0.033), l'AIC del modello finale è pari a 101.3.

Si compone così il modello finale che vede come covariate le variabili *AX*, *deltaX5*, *EIA* e *BmI* e come variabile risposta le disfunzioni delle piccole vie aeree (Tabella 4.7).

VARIABILE	Estimate	Std. Error.	z value	Pr(> Z)
INTERCETTA	-8.136	2.009	-4.049	< 0.0001
AX	4.792	0.797	6.013	< 0.0001
deltaX5	29.387	9.628	3.052	0.0023
EIA - si	1.564	0.595	2.632	0.0084
BMI	0.153	0.074	2.075	0.037

Tabella 4.7: Modello di regressione logistica dopo l'appaiamento per *over65* dato dal *Propensity Score Matching*.

Il modello finale presenta tutti i coefficienti significativi ad un livello di α fissato pari a 0.05, ha Devianza Residua pari a 91.293 con 319 gradi di libertà e un AIC di 101.29.

Interpretazione dei coefficienti

Si valuta ora il contributo di ogni variabile inserita nel modello di regressione logistica nella previsione della variabile risposta *SAD*.

- *AX* (4.79): Questo coefficiente rappresenta l'effetto di una variazione unitaria in *AX* sul log-odds della variabile di risposta *SAD*. Un aumento di una unità dell'aerea di reattanza porta a un aumento di circa 5 nel log-odds di avere disfunzioni delle piccole vie aeree.
- Il coefficiente della variabile *deltaX5* è pari a 29.39. Aumentare di una unità il valore di *deltaX5* porta a un notevole aumento di circa 30 nel log-odds della categoria di risposta positiva (*SAD* - si).
- Il coefficiente per la variabile *EIA* è pari a 1.56, dunque se un individuo presenta asma da sforzo il log-odds di avere disfunzioni delle piccole vie aeree aumenta di 1.6 unità rispetto a un individuo che non ha asma da sforzo, mantenendo costanti le altre variabili.
- *BmI* (0.153): Questo coefficiente rappresenta l'effetto di una variazione unitaria dell'indice di massa corporea sulla log-odds di *SAD*. Un aumento di una unità in *BmI* porta a un aumento di circa 0.15 nel log-odds delle disfunzioni delle piccole vie aeree.

Anche prima dell'appaiamento le variabili AX , $deltaX5$ e EIA sono risultate significative per l'inserimento nel modello, la variabile BmI invece è stata aggiunta solamente grazie a questa tecnica.

È importante considerare che i risultati sono basati sui dati e sulle unità abbinate per $over65$ dopo il *Propensity Score Matching*, il che dovrebbe aver ridotto i possibili bias presenti nei dati. Il *PSM* svolge infatti un ruolo cruciale nel processo analitico, poiché permette di migliorare la validità dell'analisi statistica e consente di ottenere stime più accurate delle relazioni tra le variabili di interesse.

4.2.3 Bontà di adattamento del modello dopo il PSM

Il valore dell'AIC del modello finale è pari a 101.29, molto minore rispetto al modello nullo di partenza, si conduce il Test del Rapporto di Verosimiglianza per confrontare i due modelli, il valore del p-value molto vicino allo zero (p-value < 0.0001) indica che il modello finale è significativamente migliore rispetto al modello iniziale.

Si costruisce dunque la tabella di corretta classificazione per valutare le prestazioni del modello viene poi calcolata l'accuratezza (Tabella 4.8)

	FALSE	TRUE
SAD = no	73	8
SAD = si	11	232

Tabella 4.8: Tabella di corretta classificazione per il modello finale dopo il *Propensity Score Matching*.

Anche dopo il *Propensity Score Matching* l'accuratezza del modello rimane pari a 0.94 circa, il che significa che le

previsioni fatte dal modello sono al 94% corrette rispetto ai dati reali.

Si costruisce la curva ROC e si calcola l'AUC per valutare le prestazioni del modello costruito (Figura 4.3).

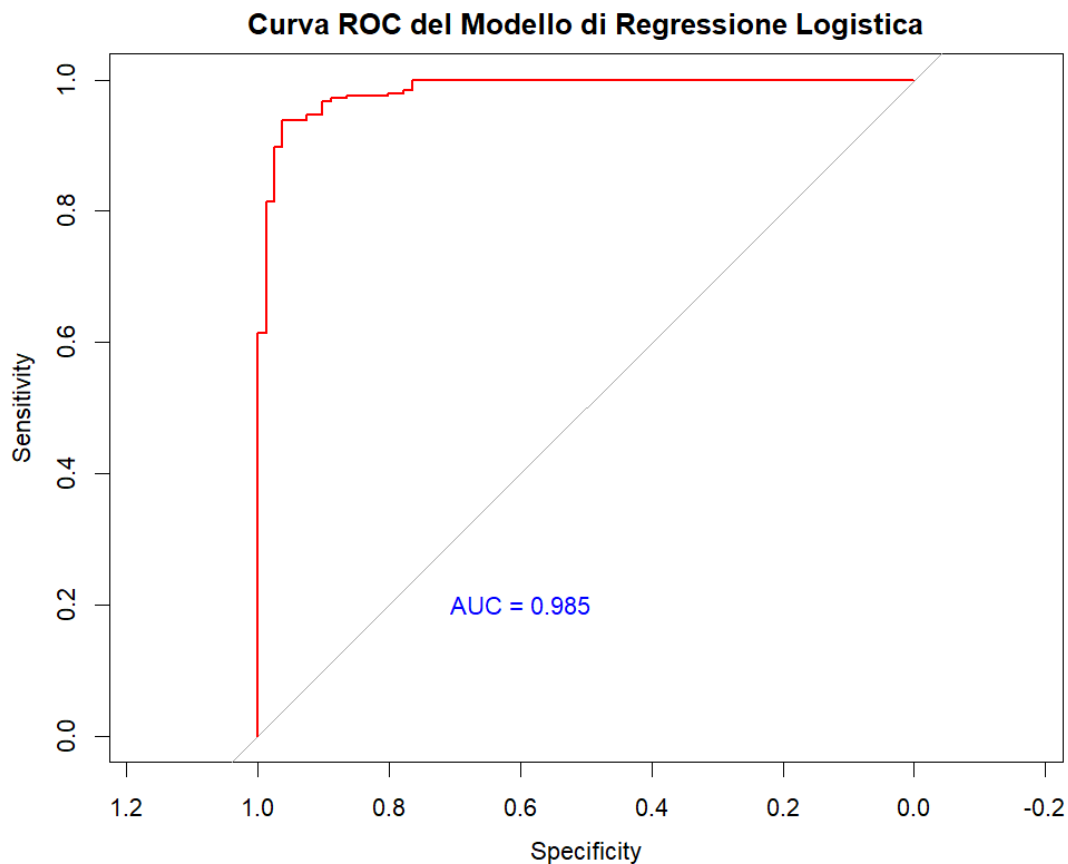


Figura 4.3: Curva ROC e AUC del modello di regressione logistica dopo il Propensity Score Matching.

La curva ROC rappresentata in Figura 4.3 si avvicina molto all'angolo in alto a sinistra indice che il modello ha una buona sensibilità.

L'AUC pari a 0.985 rappresenta l'area sottesa dalla curva ROC, in questo caso si può concludere che il modello abbia una buona capacità discriminativa.

Per concludere si conduce il test di Hosmer-Lemeshow, il valore del p-value porta a respingere l'ipotesi nulla, c'è

quindi evidenza statistica che il modello creato dopo il *Propensity Score Matching* non si adatti bene ai dati osservati ($X\text{-squared} = 324$, $df = 8$, $p\text{-value} < 0.0001$), situazione analoga al modello prima dell'appaiamento.

4.3 Considerazioni finali

Con l'utilizzo del *Propensity Score Matching* i dati hanno dimostrato che vi è stato un miglioramento del bilanciamento tra i gruppi di *over65* di trattamento e di controllo, l'unica variabile per la quale rimangono ancora alcune differenze è *duratsma*.

Si analizzano poi i dati ottenuti attraverso un'analisi bivariata e la creazione di un modello di regressione logistica.

Dai risultati ottenuti dall'analisi bivariata si può notare che le associazioni che non risultano significative con la variabile *over65* sono quelle relative a *Sesso*, *GINA*, *GINA1*, *AccessiPS*, *Ricoveri*, *Terapiaextrafine*, *fumo*, *Risvegli*, *EIA*, *duratsma*, *BmI*, *fEV1p* e *fVCp*, con un p-value maggiore di 0.05.

Prima dell'appaiamento le uniche associazioni a risultare significative con l'età (*over65*) erano quelle relative alle variabili: *SAD*, *fumo* e *duratsma*.

Il modello di regressione logistica dopo il *Propensity Score Matching* presenta tutti i coefficienti significativi.

L'analisi per la bontà di adattamento del modello basata sulla curva ROC e sull'AUC indica un buon adattamento del modello ai dati, anche se il test di H-L la smentisce, potrebbero esserci dunque dei problemi di calibrazione del modello. Il modello stimato sembra discriminare bene i dati ma non

sembra essere ben calibrato, così come per il modello prima del *Propensity Score Matching*.

Risultati e Conclusioni

Obiettivo principale di questa tesi è acquisire una maggior comprensione di quali siano i determinanti delle disfunzioni delle piccole vie aeree, valutare l'efficacia dei trattamenti somministrati il tutto con lo scopo finale di migliorare la qualità di vita dei pazienti.

Dopo una presentazione generale delle disfunzioni delle piccole vie aeree e delle caratteristiche che le contraddistinguono (Capitolo 1), è stata effettuata un'analisi esplorativa su tutte le variabili osservate per i pazienti (Capitolo 2).

Si è quindi approfondito lo studio per esaminare quali fattori abbiano maggior influenza sulla variabile dipendente *SAD* (Capitolo 3), creando un modello di regressione logistica con procedura di selezione delle variabili *forward*.

Infine si riconduce l'analisi nuovamente utilizzando il *Propensity Score Matching* in modo da ridurre i bias dei dati, per valutare eventuali differenze nei dati prima e post appaiamento.

L'analisi prima dell'appaiamento ha portato a stimare la patologia con un modello che vede interessate nella regressione logistica le cinque variabili esplicative *AX*, *EIA*, *deltaX5*, *over65* e *Riacutizzazioni*. Si ricorda che la variabile *AX* rappresenta l'area di reattanza, *deltaX5* è determinata dalla differenza tra flusso inspiratorio e espiratorio e costituiscono due dei risultati del test sull'oscillometria forzata. La

variabile *EIA* si riferisce all'asma da sforzo, *over65* all'età del paziente e *Riacutizzazioni* alla ripresa dell'andamento acuto del processo morboso.

L'analisi dopo il *Propensity Score Matching* vede interessate le variabili *AX*, *deltaX5*, *EIA* e *BmI* come covariate del modello finale. L'unica variabile aggiunta a differire dal modello prima dell'appaiamento è *BmI*. Il *Propensity Score Matching* potrebbe aver reso più probabile il rilevamento di effetti deboli o sottili che potrebbero essere stati persi prima dell'applicazione di esso.

Si può quindi concludere che, per l'analisi di *datasets* provenienti da studi clinici nei quali si possono riscontrare pazienti particolari che si distinguono da altri per caratteristiche o per evoluzione della patologia, è preferibile utilizzare il modello di regressione logistica stimato dopo il *Propensity Score Matching*, in modo da eliminare quasi completamente il bias da selezione e avere una randomizzazione dei dati.

Appendice A

Comandi R per il Propensity Score Matching

Per una maggiore comprensione del Capitolo 4, vengono riportati di seguito i comandi più significativi, che sono stati utilizzati per lo studio del *Propensity Score Matching*.

- Viene richiamata in R la “library(MatchIt)”
- Si calcolano i Propensity Score per la variabile *over65*:

```
match_obj_over <- matchit(over65 ~ fumo + BmI +  
duratsma, data = dati, family = "binomial", method =  
"nearest", distance = "glm")
```
- Si utilizza la funzione "summary" per visualizzare le statistiche di bilanciamento:

```
summary(match_obj_over)
```
- Si visualizza il bilanciamento del *dataset* con i seguenti grafici:

```
plot(match_obj_over, type = "jitter", interactive = FALSE)
```



```
plot(match_obj_over, type = "density", interactive =  
FALSE, which.xs = ~ fumo + BmI + duratsma)
```
- Si studia poi l’analisi bivariata e si crea un modello di regressione logistica attraverso i comandi:

```
matched_data_over <- match.data(match_obj_over)
View(matched_data_over)
model_matched_over <- glm(SAD ~ fumo + Bmi +
duratsma, data = matched_data_over, family = "binomial")
summary(model_matched_over)
```

- Si analizza infine la bontà di adattamento del modello attraverso la curva ROC, il valore dell'AUC e il test di Hosmer-Lemeshow.

Elenco delle figure

1.1	Panoramica sull'apparato respiratorio.	5
1.2	Rappresentazione schematica degli indici IOS rispetto alle frequenze di oscillazione: R5, R20, X5, AX e fres.	11
2.1	Boxplot delle variabili generali del paziente.	18
2.2	Boxplot delle variabili relative ai risultati del test diagnostico spirometrico.	20
2.3	Boxplot delle variabili relative ai risultati del test diagnostico basato sull'oscillometria forzata.	23
2.4	Boxplot delle variabili generali del paziente in relazione a SAD.	27
2.5	Boxplots delle variabili relative ai risultati del test diagnostico spirometrico in relazione a SAD.	29
2.6	Boxplot delle variabili relative ai risultati del test diagnostico basato sull'oscillometria forzata in relazione a SAD.	31
2.7	Boxplot delle variabili eta, fVCp, duratsma e BmI relative al tabagismo del paziente.	33
2.8	Boxplot delle relazioni tra variabili risultate significative.	35
2.9	Boxplot della variabile Dosaggio ICS distinta per GINA e della variabile Dosaggio ICS distinta per Stepterapia.	36

2.10	Correlazioni di Spearman calcolate per le variabili quantitative.	38
3.1	Curva ROC e AUC del modello di regressione logistica finale.	45
4.1	Rappresentazione dei dati dopo il <i>Propensity Score Matching</i>	50
4.2	Grafico delle densità delle variabili fumo, BmI e duratsma dopo l'applicazione del <i>Propensity Score Matching</i>	53
4.3	Curva ROC e AUC del modello di regressione logistica dopo il Propensity Score Matching. .	61

Elenco delle tabelle

2.1	Frequenze assolute e percentuali delle variabili qualitative.	16
2.2	Indicatori di sintesi della distribuzione delle variabili quantitative.	17
2.3	Risultati dei test X^2 per l'indipendenza delle variabili qualitative.	25
2.4	Mediana e scarto interquantilico delle variabili generali del paziente in relazione a SAD. . . .	27
2.5	Mediana e scarto interquantilico delle variabili relative al test spirometrico in relazione a SAD.	29
2.6	Mediana e scarto interquantilico delle variabili relative al test basato sull'oscillometria forzata in relazione a SAD.	31
2.7	Mediana e scarto interquantilico di alcune variabili in relazione al fumo.	33
3.1	Modello di regressione logistica finale.	42
3.2	Tabella di corretta classificazione per il modello finale di regressione logistica.	44
4.1	Statistiche di bilanciamento delle variabili tra il gruppo trattato e il gruppo di controllo prima dell'applicazione del <i>matching</i>	51

4.2	Statistiche di bilanciamento delle variabili tra il gruppo trattato e il gruppo di controllo dopo dell'applicazione del <i>matching</i>	51
4.3	Frequenze assolute, frequenze relative e p-value dato dal test del chi-quadro relative delle variabili qualitative distinte per over65 dopo il <i>Propensity Score Matching</i>	55
4.4	Mediana e scarto interquantilico delle variabili generali del paziente in relazione a over65 dopo il <i>Propensity Score Matching</i>	56
4.5	Mediana e scarto interquantilico delle variabili relative al test spirometrico in relazione a over65 dopo il <i>Propensity Score Matching</i>	57
4.6	Mediana e scarto interquantilico delle variabili relative al test basato sull'oscillometria forzata in relazione a over65 dopo il <i>Propensity Score Matching</i>	57
4.7	Modello di regressione logistica dopo l'appaiamento per over65 dato dal <i>Propensity Score Matching</i>	58
4.8	Tabella di corretta classificazione per il modello finale dopo il <i>Propensity Score Matching</i>	60

Bibliografia

Agresti A. (2018). *An Introduction to Categorical Data Analysis*. Wiley.

Goldman M.; Saadeh C.; Ross D. (2005). *Clinical applications of forced oscillation to assess peripheral airway function*. *Respir Physiol Neurobiol*.

Grimby G.; Takishima T.; Graham W.; Macklem P.; Mead J. (1968). *Frequency dependence of flow resistance in patients with obstructive lung disease*. *J. Clin Invest*.

Piccolo D. (1998). *Statistica*. Il Mulino.

Vineis P.; De Luca P.; Paquini P. (1987). *Numero speciale di Epidemiologia e Prevenzione*. La Nuova Italia Scientifica.

Sitografia

Boccaccino A. (2005). *L'oscillometria forzata: Pneumologia.*

<https://www.siditalia.it/diapositive>

Boccaccino A. (2018). *Spirometria e parametri di funzionalità respiratoria.*

<https://www.aprirenetwork.it/Percorso-Spirometria.pdf>

Humanitas (2023). *Enciclopedia medica.*

<https://www.humanitas.it>

Lombardo F.; Marasso R.; Valente F.; Uberti M. (2019). *Corso pratico di spirometria.*

<http://www.fimmg.org>

Ariano R. (2023). *Studio medico Dr. Renato Ariano, specialista in allergologia e pneumologia.*

<https://allergiesanremo.it/test-apparato-respiratorio>

Scardapane M. (2023). *Statistica di base degli studi real-life.*

<https://www.siditalia.it/diapositive>