

Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Triennale in
Statistica per l'Economia e l'Impresa



RELAZIONE FINALE

**ANALISI STATISTICA SU PAZIENTI TRACHEOSTOMIZZATI
AFFETTI DA *ABI***

Relatore: Prof.ssa Laura Ventura
Dipartimento di Scienze Statistiche

Laureando: Sofia Curzio
Matricola: 1102208

Anno Accademico 2016/2017

Indice

1 Introduzione allo studio	1
1.1 La tracheostomia e i pazienti con <i>ABI</i>	1
1.2 I dati	7
2 Analisi esplorative preliminari.....	11
2.1 Analisi univariate	11
2.2 Analisi bivariate	18
2.3 Conclusioni	27
3 La curva ROC.....	29
3.1 Introduzione.....	29
3.2.1 Indici sintetici	29
3.2.2 Applicazione degli indici al dataset	31
3.3.1 La curva ROC	32
3.3.2 Applicazione della curva ROC al dataset	35
4 Modelli di regressione logistica	39
4.1 I GLM e il modello logistico.....	39
4.2 Il problema di separazione	41
4.3.1 Approcci standard per la riduzione della distorsione (<i>bias</i>) della stima di massima verosimiglianza	43
4.3.2 Il metodo di Firth.....	44
4.3.3 Applicazione del metodo di Firth al dataset	47
4.3.4 La riduzione della distorsione in mediana della stima di massima verosimiglianza	52
4.3.5 Applicazione del nuovo metodo al dataset.....	55
4.4 Confronto tra i risultati ottenuti	59
5 Conclusioni	61

CAPITOLO 1

INTRODUZIONE ALLO STUDIO

1.1 La tracheostomia e i pazienti con ABI

Tracheotomia e tracheostomia sono operazioni chirurgiche effettuate con lo scopo di facilitare la respirazione del paziente nel momento in cui questa sia ostacolata o alterata. Grazie allo sviluppo di tecniche tracheostomiche sempre meno invasive, sembra essere aumentata la frequenza con cui vengono eseguiti questi tipi di intervento.

La tracheotomia è un intervento chirurgico mediante il quale viene creata un'apertura nella trachea attraverso la regione anteriore del collo. L'anastomosi della trachea con la cute del collo o la creazione di una comunicazione tra la trachea e l'ambiente mediante una cannula si definisce tracheostomia (Smeltzer *et al.*, 2010). Pertanto, a differenza di quest'ultima, nella tracheotomia non è prevista alcuna modificazione del tratto tracheale, tanto che, se non si mantiene pervia volontariamente l'apertura (mediante una cannula), questa si richiude spontaneamente nel giro di poco tempo.

Il tubo inserito nella trachea viene chiamato cannula tracheostomica; una tracheostomia può essere temporanea o definitiva (Smeltzer *et al.*, 2010). Si tratta di un procedimento elettivo, ossia effettuato sulla base di un criterio di opportunità (in quanto potrebbe essere anche differito nel tempo o sostituito da un altro trattamento), ed eseguito in anestesia locale o generale.

Mentre un'insufficienza respiratoria prolungata rappresenta probabilmente la più comune causa di tracheostomia, anche altre indicazioni, quali la diminuzione del livello di coscienza, ridotti riflessi protettivi delle vie aeree, e gravi alterazioni fisiologiche, potrebbero portare a tale intervento (Durbin, 2010).

Ci sono generalmente sei motivi per cui viene richiesta una tracheostomia:

- 1) per oltrepassare un'ostruzione delle vie aeree superiori dovuta a tumore, intervento chirurgico, trauma, presenza di corpo estraneo, oppure infezione;

- 2) per impedire danni alle vie aeree superiori o alla laringe dovuti a una prolungata intubazione translaringea;
- 3) per impedire l'aspirazione nelle vie aeree di secrezioni orali o gastriche nella persona incosciente o paralizzata (isolando la trachea dall'esofago);
- 4) per permettere un facile o frequente accesso alle vie aeree inferiori per l'aspirazione o la rimozione delle secrezioni respiratorie;
- 5) per sostituire il tubo endotracheale dopo le prime tre settimane di permanenza;
- 6) per fornire vie aeree stabili al paziente che richiede una ventilazione meccanica prolungata o un'ossigenazione di sostegno o che ha difficoltà nello svezzamento della cannula.

La ventilazione meccanica è necessaria per garantire la respirazione nel momento in cui il paziente non è in grado di respirare autonomamente (per esempio durante un intervento chirurgico in anestesia generale o durante il trattamento di gravi traumi cranici), per ossigenare il sangue quando gli sforzi ventilatori sono inadeguati e per far riposare i muscoli respiratori. Molte persone collegate a un ventilatore meccanico sono in grado di respirare spontaneamente, ma lo sforzo necessario per farlo potrebbe portarli all'esaurimento muscolare. Tramite ventilazione meccanica si può semplicemente aiutare il movimento di ventilazione, o sostituirlo completamente. Per farlo si utilizza solitamente un tubo endotracheale o inserito per mezzo di tracheotomia (Smeltzer *et al.*, 2010).

Approssimativamente il 10% dei pazienti malati critici nei reparti di terapia intensiva, richiedendo più di tre giorni di ventilazione meccanica, vengono sottoposti a tracheostomia; in tal modo è possibile fornire loro una migliore qualità di vita e ridurre i costi per l'ospedale (Zanata *et al.*, 2014). Inoltre, vi sono molti processi patologici e situazioni di emergenza che rendono necessaria una tracheostomia.

L'intervento chirurgico, di solito, viene praticato in sala operatoria o in un reparto di terapia intensiva, dove è possibile mantenere la ventilazione sotto attento controllo e garantire l'asepsi (riduzione o eliminazione di agenti contaminanti). L'incisione chirurgica viene creata tra il secondo e il terzo anello tracheale; dopo aver esposto la trachea, si inserisce una cannula tracheostomica di adeguate dimensioni dotata di cuffia. La cuffia è un manicotto gonfiabile all'estremità distale della cannula che occlude lo spazio tra le pareti della trachea e la cannula stessa, permettendo la

ventilazione meccanica senza perdite aeree e riducendo il rischio di aspirazione (passaggio di particelle di cibo nelle vie respiratorie). Oltre alla tecnica chirurgica esistono altre metodiche per confezionare una tracheostomia, come la tecnica dilatativa (Ciaglia) o quella percutanea (Fantoni), praticate soprattutto nelle terapie intensive.

La cannula da tracheostomia viene mantenuta in sede attraverso una fettuccia fissata attorno al collo della persona; si confeziona una medicazione sterile in tessuto-non tessuto posta fra la flangia della cannula e la cute, per assorbire eventuale materiale drenato e prevenire le infezioni.

Possono insorgere complicanze sia precoci sia tardive, anche anni dopo la rimozione della cannula tracheostomica. Subito dopo il confezionamento della tracheostomia, può svilupparsi un'emorragia, uno pneumotorace (presenza di aria nella cavità pleurica, cioè nello spazio che separa il polmone dalla parete toracica), un'embolia gassosa (ostruzione di un vaso sanguigno determinata dalla presenza di una bolla d'aria), l'aspirazione nelle vie aeree, un enfisema sottocutaneo o mediastinico (penetrazione di aria o di gas nelle maglie del tessuto sottocutaneo e nella parte mediana del torace, rispettivamente), una lesione del nervo laringeo ricorrente o la lesione della parete tracheale posteriore. Le complicanze a lungo termine sono l'ostruzione delle vie aeree dovuta ad accumulo di secrezioni o alla protrusione della cuffia sull'apertura distale della cannula (erniazione della cuffia), le infezioni, la rottura dell'arteria anonima, la disfagia, una fistola tracheo-esofagea, l'ischemia e la necrosi tracheale; dopo la rimozione della cannula può svilupparsi anche una stenosi tracheale (restringimento patologico della trachea) (Smeltzer *et al.*, 2010).

Da molti anni viene studiata l'interferenza della tracheostomia sulla deglutizione. La disfagia è strettamente legata alla tracheostomia, non solo perché questa procedura è indicata nei pazienti con problemi di deglutizione e di aspirazione tracheale, ma anche perché la tracheostomia in sé può causare problemi di aspirazione (Zanata *et al.*, 2014).

Disfagia è il termine medico per descrivere ogni difficoltà o fastidio legati al processo di deglutizione. La parola «disfagia», infatti, è composta dai due vocaboli «dys» (= male) e «fago» (= mangio), quindi letteralmente significa «mangio male». Una persona che

soffre di disfagia corre il rischio che il cibo, la saliva, l'acqua e le bevande possano finire parzialmente nelle vie respiratorie anziché nello stomaco. Quando ciò accade, la persona può avvertire una sensazione di soffocamento, tosse e dolore. La disfagia può essere saltuaria o persistente e, nel caso in cui il disturbo sia causato da una malattia neurodegenerativa, può aggravarsi col passare del tempo. Le variazioni nel normale processo di deglutizione correlate con l'invecchiamento sono fattori che predispongono una persona anziana a soffrire di disfagia.

In generale, le cause della disfagia sono molte: ritardi o deficit nello sviluppo della deglutizione (come avviene nelle disfagie evolutive precoci), deficit neuromotori (disfagie neurogene), asportazione di uno o più organi della regione testa/collo (disfagie post-chirurgiche). Alcune malattie neurologiche, la presenza di tumori a livello della zona testa/collo e dell'esofago, deficit metabolici e altre patologie possono contribuire alla disfagia. Non esistono prove che la disfagia sia un disturbo ereditario o con una predisposizione di tipo familiare. La disfagia è molto frequente nelle persone colpite da ictus, con stime comprese tra il 30% e il 65%.

Il passaggio di particelle di cibo nelle vie respiratorie può dare origine a infezioni polmonari che possono risultare fatali. La malnutrizione e la disidratazione sono tra i rischi più frequenti della disfagia. La persona che ne soffre, infatti, tende a limitare o rifiutare di assumere cibo e bevande per difendersi dal dolore e dal disagio causati dalla tosse come risposta al cibo che va di traverso (<https://www.deutera.com/cose-la-disfagia/>).

In alcuni casi, anche la cicatrice a seguito della tracheostomia, fissando la trachea alla pelle, può causare disfagia.

La cannula tracheostomica è un vero e proprio bypass che convoglia l'aria inspirata nel sistema respiratorio, evitando il passaggio attraverso le corde vocali. La tracheostomia impedisce il normale flusso d'aria attraverso la laringe. L'esclusione temporanea della laringe dalla funzione respiratoria ne determina però una sorta di temporaneo disuso, per quanto riguarda le sue tre funzioni essenziali: respirazione, fonazione (emissione di suoni e parole) e controllo della deglutizione. Quest'ultima, ad esempio, richiede la coordinazione tra atti respiratori e deglutitori e la corretta apertura e chiusura delle corde vocali, con movimenti finalizzati a impedire al cibo la penetrazione nelle vie

aeree. Non è pertanto difficile comprendere come la ripresa di queste tre funzioni sia spesso problematica in un paziente tracheostomizzato, ventilato, con i potenziali postumi di una prolungata intubazione e alimentato artificialmente col sondino nasogastrico che – in quanto “corpo estraneo” faringeo – di per sé costituisce un ostacolo alla corretta deglutizione.

La cannula tracheostomica riduce l'abilità di comunicare con efficacia, mentre si sa che la possibilità di parlare rappresenta un elemento fondamentale per la qualità della vita del paziente. Gli stessi familiari e operatori sanitari sono talvolta in difficoltà, in quanto l'impossibilità di esprimersi limita la realizzazione dei bisogni del paziente, situazione che diventa una delle maggiori cause di frustrazione. Nei bimbi piccoli, inoltre, la tracheostomia può compromettere l'uso della parola, in modo tale da limitarne le interazioni sociali, fondamentali per lo sviluppo delle competenze linguistiche (<http://www.uildm.org/scienza-e-medicina/approfondimenti-tematici/tracheostomia-e-comunicazione/>).

Al giorno d'oggi, si assiste ad una crescente necessità di informazioni per quanto riguarda il trattamento degli individui disfagici e tracheostomizzati e sui metodi di intervento, i quali devono essere valutati oggettivamente. L'intervento di una precoce terapia logopedica mira ad una rapida identificazione della disfagia e rappresenta un importante fattore per limitare i rischi legati all'aspirazione e per evitare complicazioni; in tal modo si può ridurre la degenza in ospedale, portando i pazienti ad una più rapida indipendenza.

Relativamente al processo di decannulazione, la maggior parte degli ospedali possiede un team interdisciplinare che si occupa di questo processo, e le interazioni tra medici, infermieri, fisioterapisti, e logopedisti riducono i tempi nell'uso della tracheostomia, accelerano lo svezzamento della cannula, e aumentano la sicurezza del paziente, riducendo i rischi in caso di fallimento. Affinché la decannulazione possa essere eseguita rapidamente ed efficacemente è necessario che il paziente si trovi in condizioni favorevoli, ossia abbia un elevato livello di coscienza, un adeguato modello di respirazione, una minima quantità di secrezioni orotracheali, una buona fonazione, una deglutizione normale, e una tosse efficace (Zanata *et al.*, 2014).

Poiché le conoscenze riguardo a come il personale sanitario decide di effettuare una decannulazione sono molto limitate, è stata condotta un'indagine con i professionisti che lavorano in questo campo, nei grandi centri del mondo, allo scopo di determinare i fattori clinici importanti nel raccomandare la decannulazione. Sono stati trovati quattro criteri determinanti per lo svezzamento della cannula tracheostomica: il livello di coscienza, la capacità di tollerare la tappatura del tubo tracheostomico, l'efficacia della tosse, e la quantità di secrezione. Di lieve importanza sono risultati essere: la comorbidità del paziente, l'eziologia dell'insufficienza respiratoria, la deglutizione, e il tasso di ossigenazione. L'età del paziente era l'unico fattore classificato come irrilevante. Questi fattori possono quindi guidare la creazione di uno specifico protocollo di decannulazione (Zanata *et al.*, 2014).

I pazienti con *Acquired Brain Injury* (*ABI*=Grave Cerebrolesione Acquisita) spesso richiedono per lunghi periodi di tempo la presenza di una tracheocannula allo scopo di proteggere le vie aeree e garantire una ventilazione meccanica prolungata. Per *ABI* si intende un danno cerebrale, dovuto a trauma cranioencefalico o ad altre cause (anossia cerebrale, emorragia, etc.), tale da determinare una condizione di coma (misurata tramite la Glasgow Coma Scale, secondo la quale il punteggio minimo è pari a 3 e indica profondo stato di incoscienza, mentre il punteggio massimo è 15; la condizione di coma si realizza nel momento in cui $GCS \leq 8$ per più di 24 ore), e menomazioni sensomotorie, cognitive o comportamentali, che comportano disabilità grave, alterando il controllo e la capacità di eseguire semplici azioni volontarie (http://www.consensusconferencegca.com/assets/files/Documento_preparatorio_CC.pdf).

La frequenza con cui viene eseguita la tracheostomia in pazienti con *ABI* contrasta con la mancanza di criteri oggettivi relativamente allo svezzamento della cannula. In letteratura, esistono diversi protocolli di decannulazione, progettati in particolare in caso di terapia intensiva acuta, dove i tubi tracheostomici dovrebbero essere rimossi dopo un breve periodo. Al contrario, si nota una mancanza di informazione sull'argomento per quanto riguarda i pazienti con *ABI*. Questi pazienti, richiedendo la cannula tracheostomica per lunghi periodi, vedono un incremento del rischio di effetti

collaterali come, ad esempio, la formazione di tessuto di granulazione, stenosi tracheale, o tracheomalacia (indebolimento delle strutture di sostegno della trachea).

È largamente accettato il concetto che una veloce e sicura decannulazione migliora i risultati, facilitando il processo di recupero del paziente, e da ciò nasce una crescente necessità di identificare un protocollo specifico per pazienti con *ABI*.

Un buon protocollo di decannulazione deve consentire di rimuovere la cannula in modo relativamente sicuro, evitando le complicanze fatali e/o la necessità di un eventuale riposizionamento.

1.2 I dati

I dati a disposizione si riferiscono ad un totale di $n=74$ pazienti con *ABI* con cannula tracheostomica. Essi sono stati reclutati subito dopo l'ammissione al Dipartimento di Neuroriabilitazione dell'Ospedale "Fondazione Ospedale San Camillo I.R.C.S.S." di Venezia e i dati sono stati rilevati tra il 2/07/2015 e il 31/07/2016.

La struttura complessa per le *ABI* dell'ospedale è destinata al ricovero di pazienti particolarmente gravi, che hanno presentato uno stato di coma di gravità alla Glasgow Coma Scale ≤ 8 per un periodo non inferiore alle 24 ore. Si tratta pertanto di pazienti provenienti da reparti di rianimazione, di neurochirurgia e di neurologia del Veneto e da fuori regione, che presentano elevate ed articolate esigenze assistenziali e riabilitative, che prevedono anche un forte coinvolgimento dei familiari.

I criteri di inclusione nel campione sono: diagnosi confermata di *ABI*, la presenza della cannula tracheostomica, l'assenza di allergia agli agrumi, età superiore ai 18 anni, stabilità clinica (intesa come assenza di febbre, sepsi o infezione attiva) e stabilità emodinamica.

Le variabili rilevate si riferiscono a caratteristiche demografiche dei pazienti e ai vari test a cui sono stati sottoposti.

Obiettivo di questo lavoro è, in primo luogo, testare le performance dei parametri quantitativi rilevati per una sicura decannulazione dei pazienti con *ABI* e, successivamente, esplorare se un possibile raggruppamento di alcuni test potrebbe risultare una migliore regola predittiva per la sicura decannulazione dei soggetti con *ABI*.

Le variabili rilevate nello studio sono le seguenti:

- Sesso: variabile dicotomica relativa al sesso del paziente (assume valore 0 se femmina e 1 se maschio);
- Età: variabile quantitativa discreta che rileva l'età del paziente in anni;
- Tempo: variabile quantitativa discreta che rileva il tempo di mantenimento della cannula in giorni;
- Tosse volontaria: variabile dicotomica relativa alla presenza e all'efficacia della tosse volontaria, misurata tramite spirometro (Pony FX Cosmed); vale 0 se il paziente registra una PCF<160 l/min (PCF=Peak Cough Flow) oppure se non riesce a tossire volontariamente a causa di deficit cognitivi (ad esempio aprassia bucco-facciale, mancanza di risposta) e 1 altrimenti;
- Tosse riflessa: variabile dicotomica relativa al risultato del test consistente nell'inalazione di un agente che provoca la tosse nebulizzato (0.1 M/L di acido citrico), utilizzando un nebulizzatore a ultrasuoni (MO-03 Norditalia Elettromedicali®); vale 1 se il paziente tossisce per almeno due volte consecutive e 0 altrimenti;
- Deglutizione: variabile dicotomica relativa alla valutazione fibro-endoscopica della deglutizione durante la somministrazione di liquidi e cibo; vale 1 in caso di un punteggio inferiore a 5 secondo la Penetration-Aspiration Scale (scala per valutare la gravità degli eventi di aspirazione) e 0 altrimenti;
- BLU: variabile dicotomica relativa ad un test per valutare le secrezioni del paziente e che consiste nel mettere 4 gocce di colorante blu dietro la lingua del paziente e nel verificare ogni 4 ore per 12 ore la presenza di secrezioni tracheali blu (vale 1 in assenza di tracce blu e 0 altrimenti);
- Chiusura: variabile dicotomica concernente l'esito del test per valutare la capacità del paziente di respirare tramite la bocca e il naso, con la cannula chiusa, per almeno 72 ore (vale 1 se il paziente supera il test e 0 altrimenti);
- Numero: variabile dicotomica che vale 1 se il numero di aspirazioni ogni 8 ore è minore o uguale a 2 e 0 altrimenti;
- Pervietà: variabile dicotomica relativa alla valutazione endoscopica della pervietà delle vie aeree (vale 1 se il diametro del lume è $\geq 50\%$ e 0 altrimenti);

- Coma: variabile dicotomica concernente il livello di coscienza del paziente misurato tramite la Glasgow Coma Scale (si assegna valore 1 se il paziente registra un punteggio ≥ 8 e 0 altrimenti);
- Ossigeno: variabile dicotomica relativa alla saturazione di ossigeno nel sangue arterioso (vale 1 se $SpO_2 > 95\%$ e 0 altrimenti);
- Esito: variabile dicotomica che rileva se, almeno in un primo momento, la decannulazione è avvenuta con successo (vale 1 in caso di successo e 0 altrimenti);
- Follow-up: variabile dicotomica riguardante l'esito della decannulazione nel paziente rilevato 48 ore dopo la rimozione della cannula tracheostomica (vale 0 in caso di fallimento della prova di passaggio a respirazione spontanea o se si verifica la necessità di una ricannulazione entro 48 ore dalla rimozione della cannula e 1 altrimenti).

Successivamente sono state costruite le seguenti tre variabili:

- Gruppo pervietà: variabile dicotomica che vale 1 quando le variabili Chiusura e Pervietà assumono entrambe valore 1 e 0 altrimenti (ossia quando almeno una delle due non assume valore 1);
- Gruppo disfagia: variabile dicotomica che vale 1 quando le variabili BLU e Deglutizione assumono entrambe valore 1 e 0 altrimenti;
- Gruppo clinico: variabile dicotomica che vale 1 quando le variabili Pervietà e Gruppo disfagia assumono entrambe valore 1 e 0 altrimenti.

CAPITOLO 2

ANALISI ESPLORATIVE PRELIMINARI

Gli obiettivi di questo capitolo sono i seguenti:

- 1) analizzare le variabili rilevate nello studio tramite strumenti di statistica descrittiva;
- 2) individuare le relazioni significative tra coppie di variabili eseguendo i test statistici considerati di volta in volta più opportuni.

Il primo punto consiste nello sviluppo di analisi univariate, mentre il secondo riguarda le analisi bivariate. Le analisi sono svolte con il software statistico RStudio (www.rstudio.com), considerando un livello di significatività del 5%.

2.1 Analisi univariate

Le analisi univariate consistono nella descrizione delle caratteristiche distributive per ogni variabile rilevata allo scopo di indentificarne comportamenti particolari. Nel dataset che si vuole analizzare sono incluse sia variabili quantitative sia variabili qualitative. Per le variabili quantitative sono riportate le seguenti misure di sintesi: minimo, primo quartile, mediana, media, terzo quartile, massimo e deviazione standard. Viene riportato anche l'intervallo di confidenza al 95% per la media. Graficamente, per intuire la struttura dei dati e controllare gli assunti possibili per l'analisi, vengono rappresentati l'istogramma con la densità stimata secondo il metodo del nucleo e il boxplot. La normalità viene valutata tramite il diagramma quantile contro quantile e il test non-parametrico di Shapiro-Wilk. Per quanto concerne le variabili qualitative dicotomiche vengono riportate le frequenze assolute, le proporzioni di successo e insuccesso e l'intervallo di confidenza al 95% per la probabilità di successo.

Le *Tabelle 1, 2 e 3* contengono i risultati di una prima analisi esplorativa univariata per le variabili quantitative e dicotomiche.

Tabella 1: Misure di sintesi e intervallo di confidenza (95%) per le variabili quantitative: età e tempo

Variabile	Minimo	Primo quartile	Mediana	Media	Terzo quartile	Massimo	Deviazione standard	Intervallo di confidenza per la media (95%)
<i>Età</i>	17.00	44.25	58.50	55.62	67.00	82.00	16.34	51.90-59.35
<i>Tempo</i>	30.00	78.50	154.00	171.60	259.80	425.00	113.28	145.80-197.42

Tabella 2: Misure di sintesi e intervallo di confidenza (95%) per le variabili dicotomiche

Variabile	Modalità	Frequenza assoluta	Frequenza relativa	Intervallo di confidenza (95%)
<i>Sesso</i>	-Maschio (1)	42	0.5676	0.4547-0.6804 (per maschio)
	-Femmina (0)	32	0.4324	
<i>Tosse volontaria</i>	-PCF $\geq$ 160 l/min (1)	20	0.2703	0.1691-0.3715 (per PCF $\geq$ 160 l/min)
	-PCF<160 l/min (0)	54	0.7297	
<i>Tosse riflessa</i>	-Tossisce almeno due volte (1)	68	0.9189	0.8567-0.9811 (per tossisce almeno due volte)
	-Tossisce meno di due volte (0)	6	0.0811	
<i>Deglutizione</i>	-PAS $\leq$ 5 (1)	55	0.7432	0.6437-0.8428 (per PAS $\leq$ 5)
	-PAS>5 (0)	19	0.2568	
<i>BLU</i>	-Assenza di tracce blu nelle secrezioni (1)	53	0.7162	0.6135-0.8189 (per assenza di tracce blu nelle secrezioni)
	-Presenza di tracce blu nelle secrezioni (0)	21	0.2838	

<i>Chiusura</i>	-Respira con la cannula chiusa per almeno 72 ore (1)	58	0.7838	0.6900-0.8776
	-Respira con la cannula chiusa per meno di 72 ore (0)	16	0.2162	(per respira con la cannula chiusa per almeno 72 ore)
<i>Numero</i>	-Numero di aspirazioni ≤ 2 (1)	56	0.7568	0.6590-0.8545
	-Numero di aspirazioni > 2 (0)	18	0.2432	(per numero di aspirazioni ≤ 2)
<i>Pervietà</i>	-Diametro del lume $\geq 50\%$ (1)	68	0.9189	0.8567-0.9811
	-Diametro del lume $< 50\%$ (0)	6	0.0811	(per diametro del lume $\geq 50\%$)
<i>Coma</i>	-GCS ≥ 8 (1)	59	0.7973	0.7057-0.8889
	-GCS < 8 (0)	15	0.2027	(per GCS ≥ 8)
<i>Ossigeno</i>	-SpO ₂ $> 95\%$ (1)	41	0.5541	0.4408-0.6673
	-SpO ₂ $\leq 95\%$ (0)	33	0.4459	(per SpO ₂ $> 95\%$)
<i>Esito</i>	-Successo della decannulazione (1)	57	0.7703	0.6744-0.8661
	-Insuccesso della decannulazione (0)	17	0.2297	(per successo della decannulazione)
<i>Follow-up</i>	-Successo della decannulazione dopo 48 ore (1)	54	0.7297	0.6285-0.8309 (per successo della decannulazione dopo 48 ore)

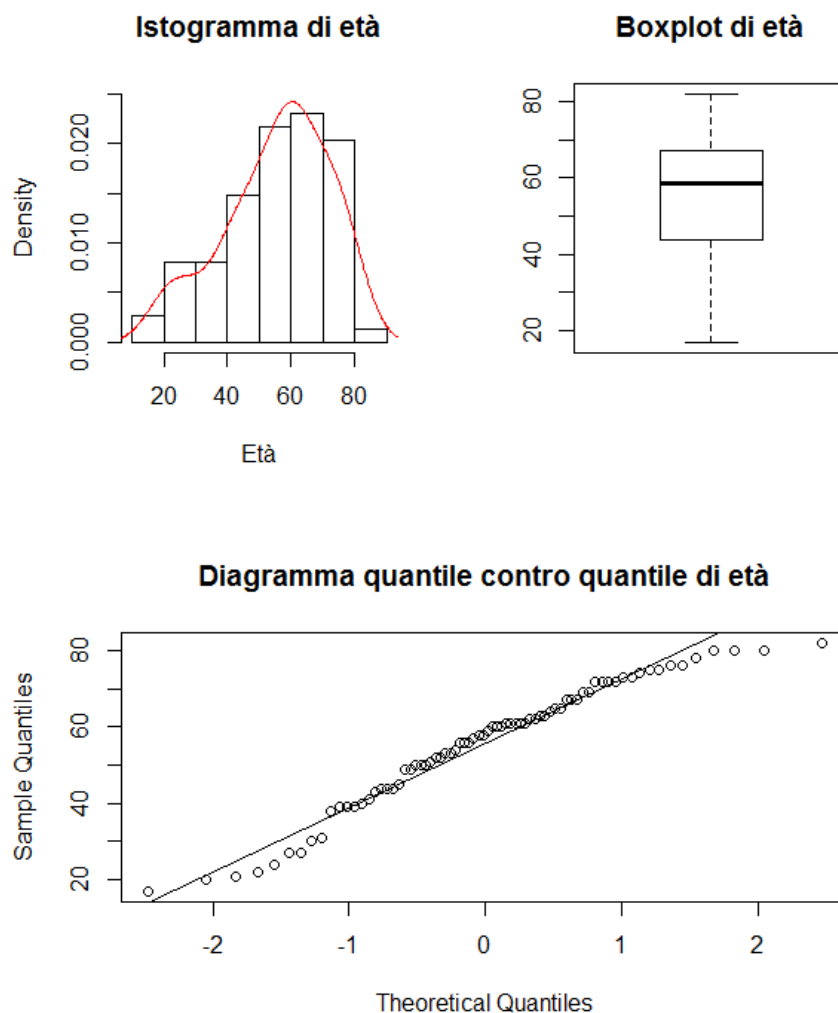
	-Insuccesso della decannulazione dopo 48 ore (0)	20	0.2703	
--	--	----	--------	--

Tabella 3: Misure di sintesi e intervallo di confidenza (95%) per le tre variabili dicotomiche costruite

Variabile	Modalità	Frequenza assoluta	Frequenza relativa	Intervallo di confidenza (95%)
<i>Gruppo pervietà</i>	-Chiusura e pervietà valgono entrambe 1 (1)	55	0.7432	0.6437-0.8428 (per chiusura e pervietà entrambe uguali a 1)
	-Chiusura e pervietà non valgono entrambe 1 (0)	19	0.2568	
<i>Gruppo disfagia</i>	-BLU e deglutizione valgono entrambe 1 (1)	48	0.6486	0.5399-0.7574 (per BLU e deglutizione entrambe uguali a 1)
	-BLU e deglutizione non valgono entrambe 1 (0)	26	0.3514	
<i>Gruppo clinico</i>	-BLU, deglutizione e pervietà assumono tutte valore 1 (1)	45	0.6081	0.4969-0.7193 (per BLU, deglutizione e pervietà tutte uguali a 1)
	-BLU, deglutizione e pervietà non assumono tutte valore 1 (0)	29	0.3919	

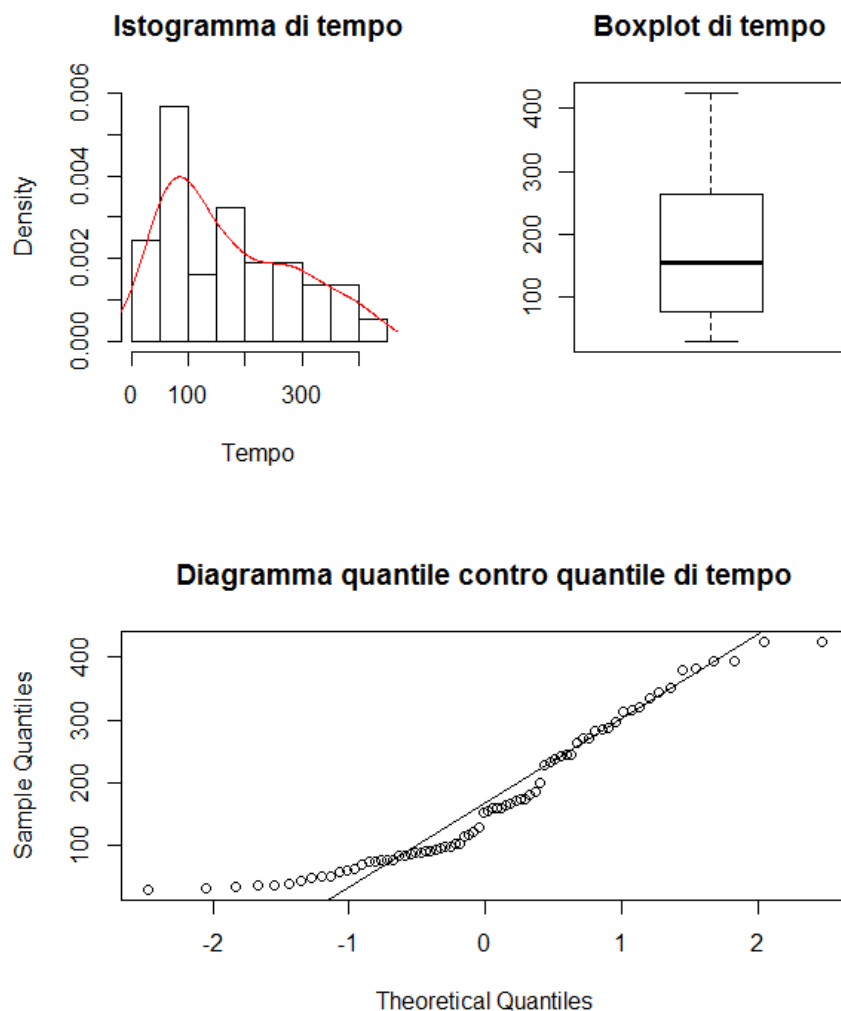
Analizzando la distribuzione della variabile quantitativa età si nota una leggera asimmetria. La distribuzione risulta spostata verso valori più elevati del range e ciò porta ad avere il baffo superiore del boxplot (*Figura 1*) più corto rispetto a quello inferiore. Dall'istogramma (*Figura 1*) emerge che la classe modale è (60,70] anni. Il diagramma quantile contro quantile (*Figura 1*) mostra un allontanamento sistematico dalla retta di riferimento sia nella coda inferiore sia in quella superiore. L'ipotesi di normalità viene rifiutata ad un livello di significatività del 5% tramite il test di Shapiro-Wilk ($SW^{OS}=0.96$, $\alpha^{OS}=0.01$).

Figura 1: Istogramma, boxplot e diagramma quantile contro quantile per la variabile età



Dalla *Figura 2* emerge che la variabile tempo di mantenimento della cannula presenta una distribuzione asimmetrica. Ciò si deduce sia dall'istogramma (*Figura 2*), che presenta come classe modale l'intervallo (50,100] giorni, sia dal boxplot (*Figura 2*), dove si nota la concentrazione della distribuzione verso valori più bassi del range. Dal diagramma quantile contro quantile (*Figura 2*) si osserva un allontanamento sistematico dalla retta di riferimento. Il test di Shapiro-Wilk rifiuta l'ipotesi di normalità ad un livello di significatività del 5% ($SW^{oss} = 0.91, \alpha^{oss} < 0.001$).

Figura 2: Iistogramma, boxplot e diagramma quantile contro quantile per la variabile tempo di mantenimento della cannula



Il campione è composto da 42 individui di sesso maschile e 32 di sesso femminile (rispettivamente il 57% e il 43% del campione).

Solo per 20 individui, i quali rappresentano il 27% del campione, si è registrato un picco di flusso della tosse (PCF) superiore a 160 l/min.

La maggior parte dei pazienti (il 92% del campione) ha tossito almeno due volte consecutive a seguito dell'inalazione di un agente che provoca la tosse, superando il test.

Gli individui che hanno superato con successo il test per la valutazione della disfagia, realizzando un punteggio pari o inferiore a 5 secondo la Penetration-Aspiration Scale, sono 55 (74%).

A seguito del Blue Dye Test, sono state rinvenute secrezioni tracheali blu in 21 pazienti (28%).

Sono 58 i pazienti (78%) che hanno superato il test consistente nella chiusura della cannula per almeno 72 ore.

I soggetti per cui si sono registrate un numero di aspirazioni pari o inferiore a 2 ogni 8 ore sono 56 (76%).

I soggetti che hanno un diametro del lume superiore al 50% sono 68 (92%).

In 59 soggetti (80%) hanno ottenuto un punteggio pari o superiore a 8 secondo la Glasgow Coma Scale per la valutazione del livello di coscienza.

In 41 (55%) hanno una saturazione di ossigeno nel sangue arterioso superiore al 95%.

La decannulazione ha avuto inizialmente esito positivo per 57 pazienti (77%), ma sono 54 (73%) i pazienti che sono riusciti a superare la prova per almeno 48 ore; ciò implica che per 3 pazienti solo inizialmente la decannulazione ha avuto successo, rendendo necessario un successivo riposizionamento della cannula.

Il gruppo pervietà è composto da 55 pazienti (74%), i quali hanno contemporaneamente sia passato il test della chiusura della cannula sia un diametro del lume superiore al 50%. Nel gruppo disfagia, invece, rientrano 48 pazienti (65%); questi hanno superato sia il Blue Dye Test sia il test per la valutazione della disfagia. Infine, 45 soggetti (61%) compongono il gruppo clinico; tali soggetti appartengono anche al gruppo disfagia e hanno un diametro del lume superiore al 50%.

2.2 Analisi bivariate

Le analisi bivariate indagano la relazione tra coppie di variabili allo scopo di individuare quelle significative. Si possono distinguere tre diversi casi.

Quando la coppia di variabili è costituita da due variabili quantitative viene riportato il diagramma di dispersione e l'indice di correlazione per misurare quanto forte è la loro relazione lineare (con relativo *p-value*). In alternativa, nel momento in cui la distribuzione delle variabili X e Y non risulta essere normale o quando le sottopopolazioni dei valori di X o Y non risultano avere la stessa varianza, si ricorre ad un indice di correlazione non parametrico dipendente solo dai ranghi delle osservazioni. In particolare, un metodo di correlazione ordinale, dovuto a Spearman, viene calcolato come segue:

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)},$$

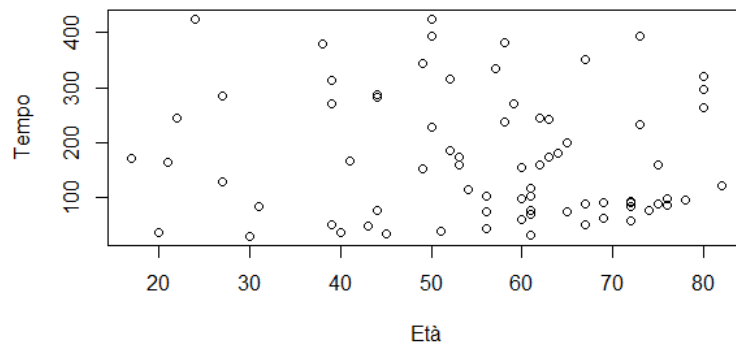
dove d_i sono le differenze tra il rango di X e di Y per l'osservazione i -esima e n il numero di osservazioni.

Quando la coppia di variabili comprende una variabile quantitativa e una qualitativa dicotomica viene, in primo luogo, verificata la normalità nei due gruppi. Tale ipotesi viene esplorata sia con i diagrammi quantile contro quantile sia conducendo il test di Shapiro-Wilk. Se quest'ipotesi viene accettata in entrambi i gruppi, l'analisi prosegue con la verifica dell'ipotesi di omoschedasticità, svolgendo il test F per la verifica di ipotesi sull'uguaglianza di due varianze. In presenza di normalità e omoschedasticità, viene condotto il test t a due campioni indipendenti. In caso di rifiuto dell'ipotesi di omoschedasticità si prosegue con il test di Welch, mentre in caso di rifiuto dell'ipotesi di normalità si prosegue con il test non-parametrico di Mann-Whitney per il confronto tra due campioni indipendenti.

Quando la coppia di variabili è costituita da due variabili dicotomiche si considera la costruzione della tabella di contingenza. Nel caso in cui tutte le celle contengano valori attesi almeno pari a 5, viene condotto il test Chi-quadro con la correzione di continuità per la verifica dell'ipotesi di indipendenza delle due variabili. In caso contrario, viene condotto il test esatto di Fisher.

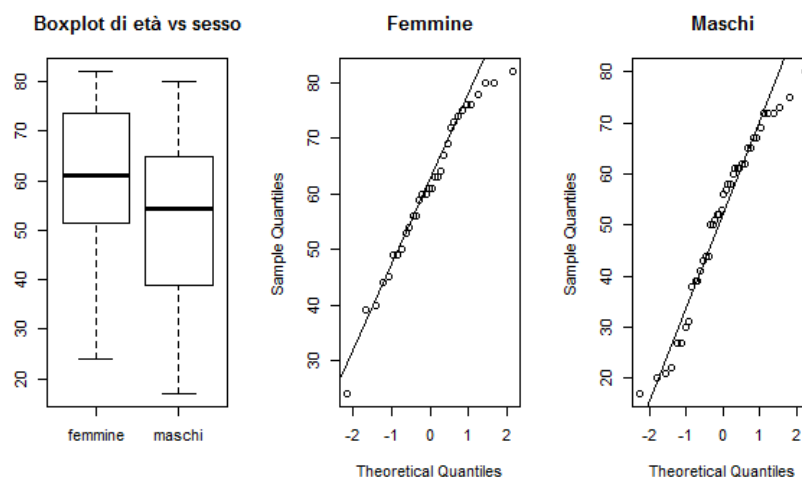
La prima coppia di variabili analizzata riguarda età e tempo di mantenimento della cannula. Dalla *Figura 3* non si nota una relazione lineare tra le due variabili. La correlazione di Spearman tra le due variabili risulta pari a -0.07 (p -value 0.53), ossia la correlazione non è significativa.

Figura 3: Diagramma di dispersione di età e tempo



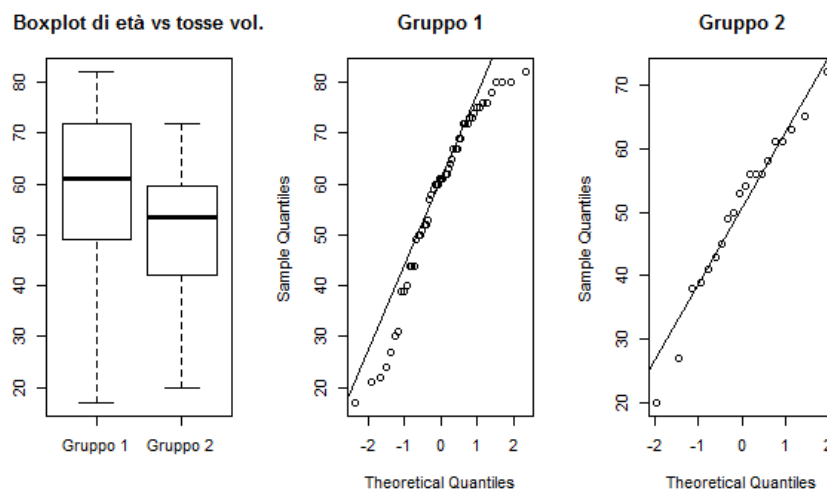
Analizzando la relazione tra le variabili età e sesso, dai diagrammi quantile contro quantile (*Figura 4*) non emergono scostamenti sistematici dalla retta di riferimento e la normalità viene accettata in entrambi i gruppi (per femmina $SW^{OSS}=0.96$ e $\alpha^{OSS}=0.33$; per maschio $SW^{OSS}=0.95$ e $\alpha^{OSS}=0.08$). L'ipotesi di omoschedasticità viene accettata ad un livello del 5% ($F^{OSS}=0.68$ e $\alpha^{OSS}=0.27$). Il test t porta a rifiutare l'ipotesi di uguaglianza delle medie nei due gruppi ($t^{OSS}=2.56$ e $\alpha^{OSS}=0.01$). L'intervallo di confidenza al 95% per la differenza delle medie tra il gruppo delle femmine e quello dei maschi è (2.11, 16.85).

Figura 4: Boxplot e diagrammi quantile contro quantile per età distintamente nel gruppo delle femmine e in quello dei maschi



Dai boxplot in *Figura 5* emerge un'età mediana diversa tra il gruppo che ha registrato un picco di flusso della tosse (PCF) inferiore a 160 l/min (gruppo 1) e quello che ha registrato un PCF pari o superiore a 160 l/min (gruppo 2). L'ipotesi di normalità viene rifiutata (*Figura 5*) con il test di Shapiro-Wilk (per gruppo 1 $SW^{OSS}=0.93$ e $\alpha^{OSS}=0.01$; per gruppo 2 $SW^{OSS}=0.96$ e $\alpha^{OSS}=0.51$). Il test di Mann-Whitney porta al rifiuto dell'ipotesi di uguaglianza delle mediane nei due gruppi ($W^{OSS}=716.5$ e $\alpha^{OSS}=0.03$). La mediana è significativamente superiore nel gruppo 1; l'intervallo di confidenza al 95% per la differenza delle mediane tra il gruppo 1 e il gruppo 2 è (0.999, 16).

Figura 5: Boxplot e diagrammi quantile contro quantile per età distintamente nei due gruppi di tosse volontaria



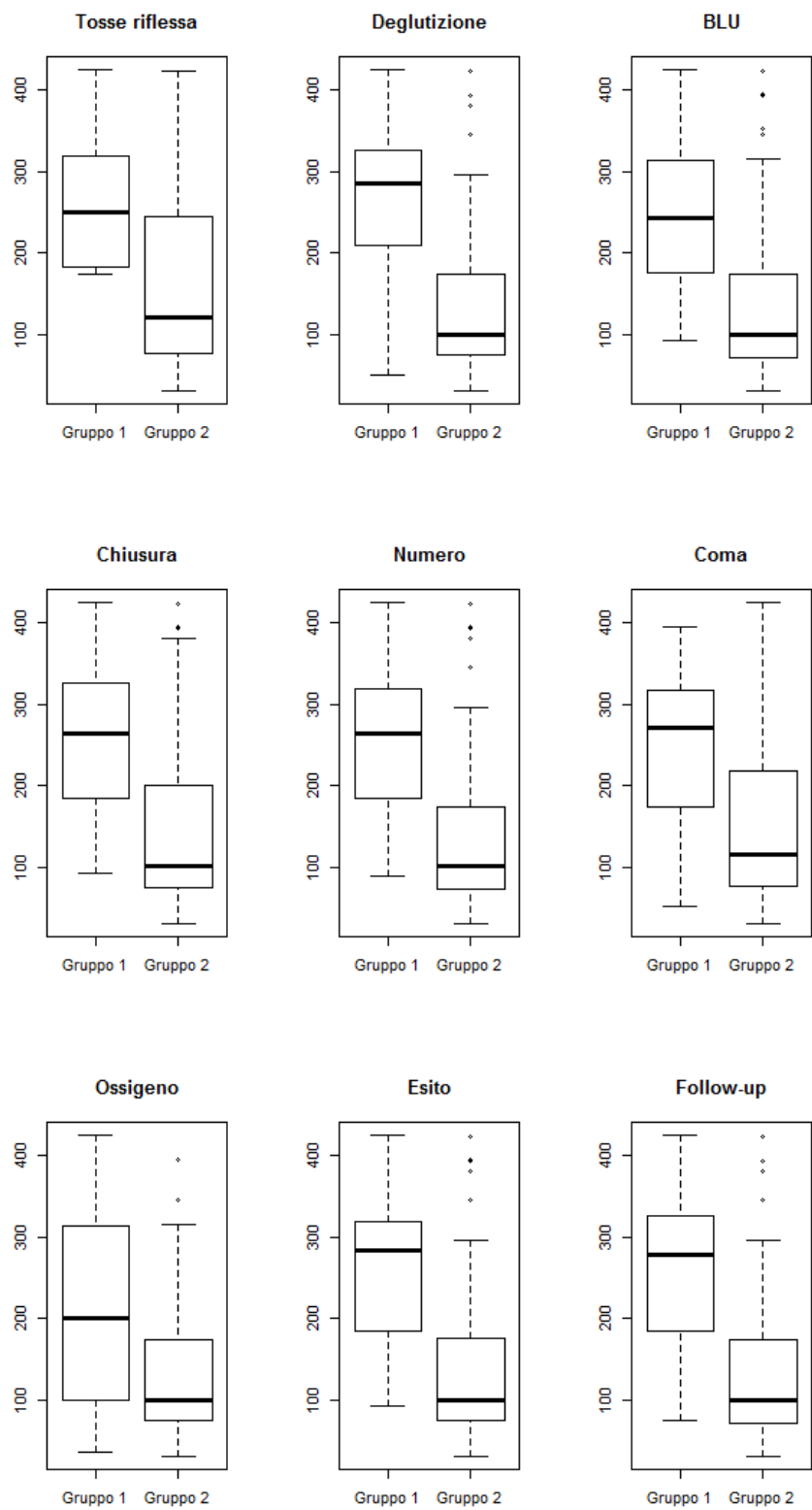
Tutti gli altri confronti tra età e le rimanenti variabili dicotomiche sono risultati non significativi ($\alpha^{OSS}>0.1$).

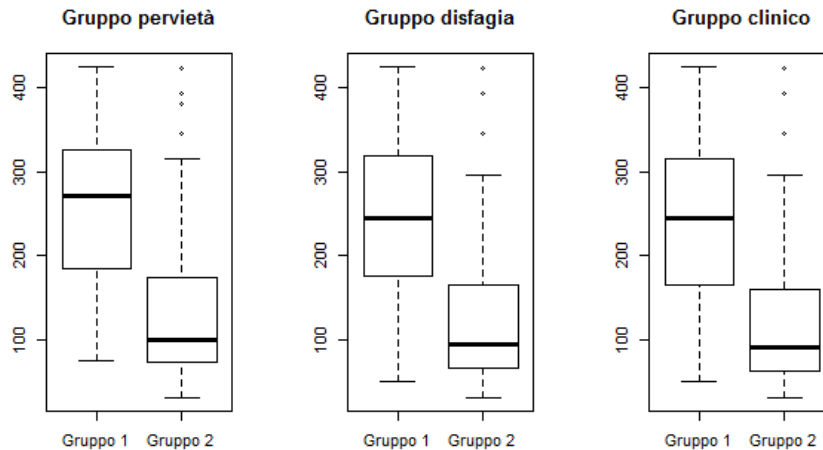
La variabile tempo di mantenimento della cannula ha una relazione significativa con la maggior parte delle variabili dicotomiche rilevate. La *Tabella 4* riporta, per ogni confronto risultato significativo, i valori assunti dai test e i relativi *p-values*. Il gruppo 1 è quello in cui la variabile dicotomica considerata assume valore 0, mentre il gruppo 2 è quello in cui la stessa assume valore 1. Come si può notare, la normalità è stata accettata sempre in uno solo dei due gruppi e ciò ha portato a svolgere tutti test di Mann-Whitney non-parametrici. L'intervallo di confidenza al 95% per la differenza delle mediane ha portato ad evidenziare una mediana significativamente superiore nel gruppo 1 per tutti i confronti, come indicano i boxplot rappresentati in *Figura 6*.

Tabella 4: Valori osservati dei test e relativi p-values per tempo di mantenimento della cannula e alcune variabili dicotomiche

Variabile: Tempo	Gruppo 1	Gruppo 2	W^{OSS}	p-value
<i>Tosse riflessa</i>	$SW^{OSS}=0.85$ $\alpha^{OSS}=0.16$	$SW^{OSS}=0.90$ $\alpha^{OSS}<0.001$	322	0.02
<i>Deglutizione</i>	$SW^{OSS}=0.96$ $\alpha^{OSS}=0.66$	$SW^{OSS}=0.86$ $\alpha^{OSS}<0.001$	860.5	<0.001
<i>BLU</i>	$SW^{OSS}=0.97$ $\alpha^{OSS}=0.67$	$SW^{OSS}=0.84$ $\alpha^{OSS}<0.001$	878.5	<0.001
<i>Chiusura</i>	$SW^{OSS}=0.98$ $\alpha^{OSS}=0.96$	$SW^{OSS}=0.87$ $\alpha^{OSS}<0.001$	742.5	<0.001
<i>Numero</i>	$SW^{OSS}=0.97$ $\alpha^{OSS}=0.85$	$SW^{OSS}=0.86$ $\alpha^{OSS}<0.001$	811.5	<0.001
<i>Coma</i>	$SW^{OSS}=0.93$ $\alpha^{OSS}=0.24$	$SW^{OSS}=0.88$ $\alpha^{OSS}<0.001$	625	0.01
<i>Ossigeno</i>	$SW^{OSS}=0.94$ $\alpha^{OSS}=0.06$	$SW^{OSS}=0.89$ $\alpha^{OSS}<0.001$	919	0.01
<i>Esito</i>	$SW^{OSS}=0.98$ $\alpha^{OSS}=0.91$	$SW^{OSS}=0.86$ $\alpha^{OSS}<0.001$	789.5	<0.001
<i>Follow-up</i>	$SW^{OSS}=0.97$ $\alpha^{OSS}=0.70$	$SW^{OSS}=0.86$ $\alpha^{OSS}<0.001$	877	<0.001
<i>Gruppo pervietà</i>	$SW^{OSS}=0.97$ $\alpha^{OSS}=0.79$	$SW^{OSS}=0.87$ $\alpha^{OSS}<0.001$	833	<0.001
<i>Gruppo disfagia</i>	$SW^{OSS}=0.98$ $\alpha^{OSS}=0.76$	$SW^{OSS}=0.83$ $\alpha^{OSS}<0.001$	1025	<0.001
<i>Gruppo clinico</i>	$SW^{OSS}=0.97$ $\alpha^{OSS}=0.66$	$SW^{OSS}=0.82$ $\alpha^{OSS}<0.001$	1046.5	<0.001

Figura 6: Boxplot di tempo di mantenimento della cannula e le variabili dicotomiche tosse riflessa, deglutizione, BLU, chiusura, numero di aspirazioni, coma, ossigeno, esito, follow-up e le tre variabili artificiali





Le analisi successive coinvolgono coppie di variabili dicotomiche.

La variabile sesso è stocasticamente indipendente da tutte le altre variabili dicotomiche, come anche la variabile tosse volontaria ($\alpha^{oss} > 0.1$).

La *Tabella 5* contiene tutte le coppie di variabili dicotomiche dipendenti e riepiloga i risultati del test Chi-quadro o, in alternativa, del test esatto di Fisher. L'ultima colonna riporta i valori del rapporto tra quote di due eventi (*odds ratio*). In particolare, il rapporto tra quote viene impiegato per confrontare la probabilità di uno stesso evento in diverse circostanze. Pertanto il rapporto delle quote è una misura di quanto maggiore (o minore) siano gli *odds* di manifestare un certo evento per quei soggetti che posseggono un determinato fattore di rischio.

Ad esempio, se A è l'evento per cui il paziente tossisce meno di due volte a seguito dell'inalazione di un agente che provoca la tosse e $P(A|P < 5)$ e $P(A|P \geq 5)$ sono le probabilità di tossire meno di due volte in un paziente che ha ottenuto un punteggio $PAS < 5$ e in un paziente che ha ottenuto un punteggio $PAS \geq 5$, rispettivamente, il rapporto delle quote è:

$$OR = \frac{o_{A|P < 5}}{o_{A|P \geq 5}} = \frac{\frac{P(A|P < 5)}{1 - P(A|P < 5)}}{\frac{P(A|P \geq 5)}{1 - P(A|P \geq 5)}}$$

e rappresenta la misura in cui l'evento condizionante ($PAS \geq 5$) aumenta la quota (e quindi la probabilità) di tossire meno di due volte. Un rapporto di quote unitario è indice del fatto che l'evento condizionante non ha influenza sulla probabilità d'interesse.

Tabella 5: Test Chi-quadro con relativo p-value o p-value del test esatto di Fisher per tutte le possibili coppie di variabili dicotomiche risultate significative

Coppia di variabili	$\chi^2_{\text{c}}^{\text{oss}}$	p-value del test Chi- quadro	p-value del test di Fisher	Odds ratio
<i>Tosse riflessa-Deglutizione</i>	-	-	0.004	18.31
<i>Tosse riflessa-BLU</i>	-	-	0.05	5.83
<i>Tosse riflessa-Chiusura</i>	-	-	0.02	8.94
<i>Tosse riflessa-Numero</i>	-	-	0.003	20.00
<i>Tosse riflessa-Esito</i>	-	-	0.002	21.96
<i>Tosse riflessa-Follow-up</i>	-	-	0.005	16.83
<i>Tosse riflessa-Gruppo pervietà</i>	-	-	0.03	6.83
<i>Tosse riflessa-Gruppo disfagia</i>	-	-	0.02	10.82
<i>Tosse riflessa-Gruppo clinico</i>	-	-	0.03	8.90
<i>Deglutizione-BLU</i>	22.91	<0.001	-	18.04
<i>Deglutizione-Chiusura</i>	-	-	<0.001	169.20
<i>Deglutizione-Numero</i>	-	-	<0.001	57.49
<i>Deglutizione-Coma</i>	-	-	0.002	7.09
<i>Deglutizione-Ossigeno</i>	14.15	<0.001	-	11.49
<i>Deglutizione-Esito</i>	-	-	<0.001	229.66
<i>Deglutizione-Follow-up</i>	46.38	<0.001	-	122.9
<i>Deglutizione-Gruppo pervietà</i>	-	-	<0.001	79.65
<i>Deglutizione-Gruppo disfagia</i>	43.44	<0.001	-	-
<i>Deglutizione-Gruppo clinico</i>	36.31	<0.001	-	-
<i>BLU-Chiusura</i>	-	-	<0.001	25.20
<i>BLU-Numero</i>	31.86	<0.001	-	38.04
<i>BLU-Coma</i>	-	-	0.02	3.95
<i>BLU-Ossigeno</i>	10.13	0.001	-	6.59
<i>BLU-Esito</i>	-	-	<0.001	18.70
<i>BLU-Follow-up</i>	15.7	<0.001	-	10.22
<i>BLU-Gruppo pervietà</i>	17.60	<0.001	-	12.12

<i>BLU-Gruppo disfagia</i>	50.23	<0.001	-	-
<i>BLU-Gruppo clinico</i>	42.01	<0.001	-	-
<i>Chiusura-Numero</i>	-	-	<0.001	41.41
<i>Chiusura-Coma</i>	-	-	0.002	7.02
<i>Chiusura-Ossigeno</i>	13.08	<0.001	-	13.84
<i>Chiusura-Esito</i>	-	-	<0.001	-
<i>Chiusura-Follow-up</i>	-	-	<0.001	-
<i>Chiusura-Gruppo pervietà</i>	-	-	<0.001	-
<i>Chiusura-Gruppo disfagia</i>	27.58	<0.001	-	59.53
<i>Chiusura-Gruppo clinico</i>	28.51	<0.001	-	-
<i>Numero-Coma</i>	-	-	0.04	3.73
<i>Numero-Ossigeno</i>	16.59	<0.001	-	17.59
<i>Numero-Esito</i>	-	-	<0.001	54.73
<i>Numero-Follow-up</i>	-	-	<0.001	26.94
<i>Numero-Gruppo pervietà</i>	-	-	<0.001	20.22
<i>Numero-Gruppo disfagia</i>	33.35	<0.001	-	81.20
<i>Numero-Gruppo clinico</i>	27.49	<0.001	-	58.20
<i>Pervietà-Coma</i>	-	-	0.014	9.88
<i>Pervietà-Follow-up</i>	-	-	<0.001	-
<i>Pervietà-Gruppo pervietà</i>	-	-	<0.001	-
<i>Pervietà-Gruppo clinico</i>	-	-	0.003	-
<i>Coma-Esito</i>	-	-	0.004	6.14
<i>Coma-Follow-up</i>	-	-	<0.001	14.46
<i>Coma-Gruppo pervietà</i>	-	-	<0.001	16.50
<i>Coma-Gruppo disfagia</i>	3.83	0.05	-	3.63
<i>Coma-Gruppo clinico</i>	7.49	0.006	-	6.09
<i>Ossigeno-Esito</i>	10.83	<0.001	-	9.04
<i>Ossigeno-Follow-up</i>	8.64	0.003	-	5.84
<i>Ossigeno-Gruppo pervietà</i>	10.41	0.001	-	7.48
<i>Ossigeno-Gruppo disfagia</i>	11.44	<0.001	-	6.40
<i>Ossigeno-Gruppo clinico</i>	9.90	0.002	-	5.33

<i>Esito-Follow-up</i>	-	-	<0.001	-
<i>Esito-Gruppo pervietà</i>	-	-	<0.001	229.66
<i>Esito-Gruppo disfagia</i>	30.41	<0.001	-	69.36
<i>Esito-Gruppo clinico</i>	31.01	<0.001	-	-
<i>Followup-Gruppo pervietà</i>	64.13	<0.001	-	-
<i>Followup-Gruppo disfagia</i>	26.98	<0.001	-	26.48
<i>Followup-Gruppo clinico</i>	39.10	<0.001	-	-
<i>Gruppo pervietà-Gruppo disfagia</i>	24.20	<0.001	-	22.57
<i>Gruppo pervietà-Gruppo clinico</i>	36.31	<0.001	-	-
<i>Gruppo disfagia-Gruppo clinico</i>	58.33	<0.001	-	-

Ricordando che il gruppo pervietà è composto dai soggetti che hanno superato il test della chiusura della cannula e che hanno un diametro del lume superiore o uguale al 50%, che il gruppo disfagia è composto da coloro che hanno superato il Blue Dye Test e il test per la valutazione della disfagia e che il gruppo clinico è composto dai soggetti inclusi nel gruppo disfagia e che hanno anche un diametro del lume superiore al 50%, emergono alcune importanti osservazioni:

- tutti i pazienti che non hanno superato il test della chiusura della cannula tracheostomica per almeno 72 ore non hanno superato neanche la decannulazione temporanea e, quindi, neppure quella definitiva e non sono stati ammessi al gruppo clinico;
- tutti i pazienti che hanno un diametro del lume inferiore al 50% non hanno superato la decannulazione definitiva;
- tutti i pazienti che non hanno superato la decannulazione temporanea (e quindi quella definitiva) non sono stati ammessi al gruppo clinico;
- tutti i pazienti che hanno superato la prova di decannulazione definitiva sono stati ammessi al gruppo pervietà;
- tutti i pazienti che non sono rientrati nel gruppo pervietà non sono rientrati nemmeno nel gruppo clinico.

2.3 Conclusioni

Dalle analisi esplorative presentate in questo capitolo è emerso che le variabili quantitative età e tempo di mantenimento della cannula non hanno una distribuzione normale, in quanto presentano una distribuzione asimmetrica, e non sono tra loro correlate. Tranne che per la variabile tosse volontaria, per tutte le altre variabili dicotomiche che riguardano i test è più elevata la proporzione di pazienti che ottiene un risultato positivo (che indica il superamento del test) rispetto a quella dei pazienti con risultato negativo.

Nello sviluppo di una possibile analisi di regressione sarà importante tenere presente tutte le dipendenze tra variabili contenute nelle *Tabelle 4* e *5*, nonché la relazione tra età e sesso e tra età e tosse volontaria. Va sottolineato, infine, che la variabile follow-up (per la quale si può nutrire un interesse di tipo previsivo) è dipendente da quasi tutte le variabili rilevate (è stocasticamente indipendente solo da età, sesso e tosse volontaria).

CAPITOLO 3

LA CURVA ROC

3.1 Introduzione

I test diagnostici ideali (i cosiddetti *golden standard*) discriminano perfettamente i malati dai sani, cioè, una volta eseguito il test, gli individui vengono classificati con assoluta certezza come affetti o non affetti dalla malattia di interesse. In genere è difficile che un determinato test diagnostico discrimini in maniera netta i malati dai sani; ciò implica che le distribuzioni dei risultati del test sono parzialmente sovrapposte negli individui affetti e non affetti da una specifica malattia e risulta necessario il calcolo del grado di incertezza della classificazione. Se il risultato del test diagnostico di interesse è una variabile binaria (affetto/non affetto), si calcolano la sensibilità, la specificità, il potere predittivo positivo, il potere predittivo negativo e l'accuratezza. Se invece il risultato del test è una variabile continua, si utilizza l'analisi della curva ROC. La curva ROC è una tecnica statistica che misura l'accuratezza di un test diagnostico lungo tutto il dominio della variabile criterio scelta. Poiché la curva ROC misura l'accordo tra il test di interesse e la presenza/assenza di una specifica malattia (così come identificata da un *golden standard*), essa rappresenta il metodo d'elezione per validare un test diagnostico. La curva ROC permette anche di identificare il valore soglia ottimale (il cosiddetto *best cut-off*) (D'Arrigo *et al.*, 2011).

3.2.1 Indici sintetici

Il potere diagnostico di un test include la sensibilità, la specificità, il potere predittivo positivo, il potere predittivo negativo e l'accuratezza. La *Tabella 6* mette in rapporto, in termini generali, i risultati di un ipotetico test diagnostico (che può essere positivo o negativo) con la presenza/assenza di una specifica malattia.

Tabella 6: Matrice di confusione

	Paziente malato	Paziente sano	Totale
Test positivo	TP	FP	TP+FP
Test negativo	FN	TN	FN+TN
Totale	TP+FN	FP+TN	

Nella *Tabella 6* (denominata *matrice di confusione* o *tabella di errata classificazione*) i pazienti sono stati suddivisi in quattro categorie, ossia:

- Veri Positivi (TP=True Positive): pazienti malati correttamente classificati come positivi;
- Falsi Negativi (FN=False Negative): pazienti malati erroneamente classificati come negativi;
- Veri Negativi (TN=True Negative): pazienti sani correttamente classificati come negativi;
- Falsi Positivi (FP=False Positive): pazienti sani erroneamente classificati come positivi.

Un test è positivo quando mette in luce la presenza della malattia.

Utilizzando i dati riportati in ciascuna cella della *Tabella 6* è possibile calcolare i cinque indici sintetici che esprimono la qualità della classificazione, cioè:

- la sensibilità: la proporzione di pazienti con test positivo tra tutti quelli che hanno la malattia $[TP/(TP+FN)]$, cioè la proporzione di Veri Positivi rispetto al numero totale di positivi effettivi;
- la specificità: la proporzione di pazienti con test negativo tra tutti quelli che sono sani $[TN/(FP+TN)]$, cioè la proporzione di Veri Negativi rispetto al numero totale di negativi effettivi;
- il potere predittivo positivo: la proporzione di pazienti malati tra tutti quelli che sono positivi al test $[TP/(TP+FP)]$;
- il potere predittivo negativo: la proporzione di pazienti sani tra tutti quelli che sono negativi al test $[TN/(FN+TN)]$;
- l'accuratezza: la proporzione di pazienti correttamente classificati $[(TP+TN)/(TP+FN+FP+TN)]$.

Un test diagnostico è sensibile al 100% quando tutti i malati risultano positivi ed è specifico al 100% quando tutti i sani risultano negativi.

La sensibilità e la specificità sono misure indipendenti dalla prevalenza della malattia, cioè non sono influenzate dalla frequenza con cui una certa patologia è presente in uno specifico campione di individui (prevalenza: $[(TP+FN)/(TP+FN+FP+TN)]$). Viceversa, il potere predittivo positivo e negativo sono strettamente dipendenti dalla frequenza della malattia di interesse.

Si noti che se un test ha un'ottima specificità, allora è basso il rischio di Falsi Positivi (pazienti che pur presentando valori anomali non sono affetti dalla patologia); analogamente, se un test ha un'ottima sensibilità, allora è basso il rischio di Falsi Negativi (pazienti che pur presentando valori normali sono affetti dalla patologia).

Nella scelta tra più test si privilegerà quello con sensibilità maggiore o con specificità maggiore, a seconda che l'obiettivo sia, rispettivamente, individuare il maggiore numero di malati o il maggiore numero di sani.

3.2.2 Applicazione degli indici al dataset

La variabile follow-up descrive l'esito della rimozione della cannula dopo 48 ore ed i pazienti che registrano un valore pari a 1 superano la prova, mentre i pazienti che registrano un valore pari a 0 non superano la prova. Questa variabile rappresenta il *golden test*, ossia il test diagnostico più accurato per la valutazione del processo di decannulazione. Nella *Tabella 7* vengono riportati gli indici sintetici precedentemente descritti tra la variabile follow-up e tutti i test dicotomici che sono stati eseguiti e che hanno una relazione significativa con il *golden test* (sono state incluse anche le tre variabili ottenute come combinazione di più test, i gruppi).

Tabella 7: Indici sintetici

Follow-up	Sensibilità	Specificità	Potere Predittivo Positivo	Potere Predittivo Negativo	Accuratezza
<i>Tosse riflessa</i>	0.25	0.98	0.83	0.78	0.78
<i>Deglutizione</i>	0.85	0.96	0.89	0.95	0.93
<i>BLU</i>	0.65	0.85	0.62	0.87	0.80

<i>Chiusura</i>	0.80	1	1	0.93	0.95
<i>Numero</i>	0.70	0.93	0.78	0.89	0.86
<i>Pervietà</i>	0.30	1	1	0.79	0.81
<i>Coma</i>	0.55	0.93	0.73	0.85	0.82
<i>Ossigeno</i>	0.75	0.67	0.45	0.88	0.69
<i>Esito</i>	0.85	1	1	0.95	0.96
<i>Gruppo pervietà</i>	0.95	1	1	0.98	0.99
<i>Gruppo disfagia</i>	0.85	0.83	0.65	0.94	0.84
<i>Gruppo clinico</i>	1	0.83	0.69	1	0.88

Se si vuole privilegiare la sensibilità, il test diagnostico migliore risulta essere il gruppo clinico, mentre, nel caso in cui si voglia privilegiare la specificità, si presentano diverse alternative per quanto riguarda la scelta del test migliore: chiusura, pervietà, esito e gruppo pervietà presentano tutte una specificità del 100%. Visti i problemi che può comportare il mantenimento della cannula nella trachea per lunghi periodi di tempo, probabilmente si preferirà riuscire a classificare correttamente il maggior numero di pazienti che superano il test di decannulazione, ossia si preferirà il test con maggiore specificità.

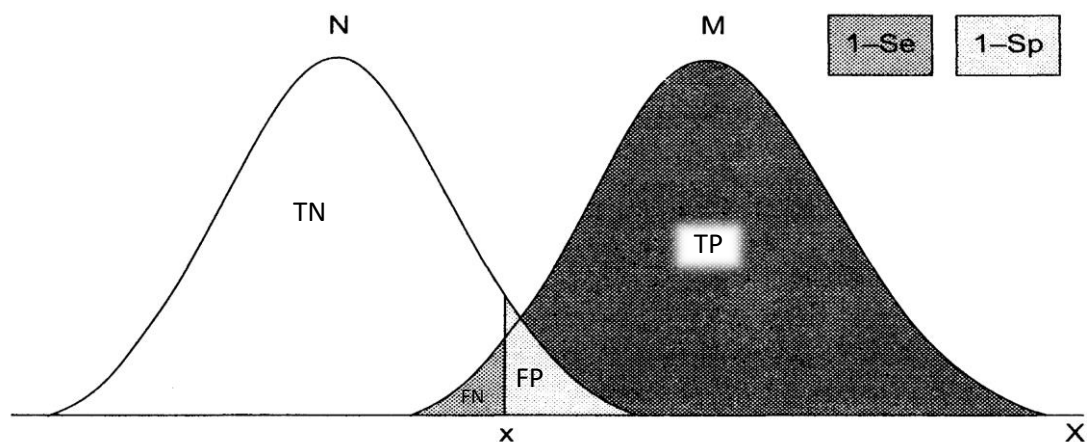
3.3.1 La curva ROC

L'analisi della curva ROC è un metodo che sempre più va diffondendosi in ambito biomedico e consente di esprimere, rappresentare e valutare l'adeguatezza di un test diagnostico misurato su scala continua. Tale metodo si basa sulla teoria statistica delle decisioni e, dalla metà degli anni '40, venne sviluppato nel contesto dell'identificazione dei segnali elettronici. ROC è l'acronimo dell'espressione inglese *"Receiver Operating Characteristic"* (*Curva Operativa Caratteristica*). Le prime applicazioni al contesto clinico risalgono alla metà degli anni '70 (Duca, 1995).

I valori di sensibilità e specificità, come altre misure relative alla performance della classificazione, dipendono dal valore soglia usato per classificare un test come positivo. A titolo esemplificativo si può considerare il caso di un test diagnostico basato sulla distribuzione di una variabile quantitativa (X) nella popolazione dei soggetti non patologici (N) ed in quella dei soggetti patologici (M). Nel caso ipotizzato, i soggetti patologici, ossia quelli estratti da M , manifestano valori generalmente superiori rispetto ai soggetti non patologici (estratti da N).

La *Figura 7* schematizza la situazione di una variabile criterio distribuita in N ed in M in forma gaussiana ed è evidente che i valori di sensibilità e specificità variano al variare del valore della X scelto come soglia. La bontà della discriminazione fra soggetti estratti da N e soggetti estratti da M dipende solo dall'entità della sovrapposizione fra le due distribuzioni (Duca, 1995).

Figura 7: Distribuzione di una variabile criterio per un test diagnostico (X) nella popolazione dei soggetti patologici (M) e non patologici (N). In corrispondenza di un arbitrario valore soglia si individuano le quattro categorie di pazienti (TN, TP, FN, FP).



La curva ROC è utile in particolare per il confronto tra due distribuzioni continue e viene costruita considerando tutti i possibili valori del test, per ognuno dei quali si calcola la proporzione di Veri Positivi (la sensibilità) e la proporzione di Falsi Positivi, ottenuta come complemento a uno della Specificità. Successivamente si congiungono tutti i punti di coordinate (1-Specificità, Sensibilità).

Se un ipotetico nuovo test discriminasse perfettamente i malati dai sani, sarebbe rappresentato da una curva ROC che passa per l'angolo superiore sinistro del sistema di assi cartesiani (100% Specificità, 100% Sensibilità). In caso contrario, ossia se il

nuovo test non discriminasse per niente i malati dai sani, la curva ROC coinciderebbe con la bisettrice dell'angolo nell'origine e ciò corrisponde alla classificazione casuale dei pazienti. È immediata l'estensione dell'utilizzo della curva ROC nell'ambito del confronto dell'efficienza tra test diagnostici differenti di una particolare patologia: la curva più spostata verso il centro del grafico appartiene all'esame peggiore.

Una migliore e più completa descrizione dell'accuratezza della classificazione è rappresentata dall'area sottesa alla curva ROC, indice che può risultare di particolare interesse anche in contesti diversi da quello diagnostico (come ad esempio nell'analisi della sopravvivenza o in uno studio caso-controllo in cui l'esposizione sia espressa da una variabile quantitativa).

L'area sottesa alla curva ROC (*AUC*, acronimo dei termini inglesi "*Area Under the Curve*"), in genere, si calcola per avere una misura sintetica ed oggettiva dell'efficienza di un test.

Riprendendo la *Figura 7*, sia *N* la variabile che rappresenta la misura nel gruppo dei pazienti sani e *M* quella nel gruppo dei malati. L'*AUC* può essere espressa come:

$$AUC=P(N<M).$$

L'area sotto la curva può assumere valori compresi tra 0.5 e 1.0 e al crescere dell'area sotto la curva aumenta il potere discriminante del test. Nella realtà, si considera adeguato un test diagnostico con un'area sotto la curva $\geq 80\%$.

Per l'interpretazione dei valori dell'area sottesa alla curva ROC è possibile riferirsi alla classificazione proposta da Swets (1988):

- $AUC=0.5$: il test non è informativo;
- $0.5 < AUC \leq 0.7$: il test è poco accurato;
- $0.7 < AUC \leq 0.9$: il test è moderatamente accurato;
- $0.9 < AUC < 1.0$: il test è altamente accurato;
- $AUC=1$: il test è perfetto.

L'area sotto la curva può essere calcolata seguendo diversi approcci, tra cui: geometrico, applicando la regola del trapezoide; parametrico, assumendo che sottese al processo di classificazione siano due distribuzioni gaussiane; non parametrico, calcolando la statistica W di Wilcoxon.

Inoltre espressioni teoriche per l'AUC sono disponibili con riferimento a diverse assunzioni distributive su N e M.

Per quanto riguarda la scelta della soglia per discriminare tra i due gruppi di pazienti, questa può essere determinata in maniera automatica secondo un criterio di ottimalità. Ne sono degli esempi:

- l'accuratezza: si sceglie come soglia il valore che permette di classificare correttamente la maggior parte dei pazienti;
- il criterio della minima distanza dal punto ottimale (0,1), calcolato come:

$$D_{(0,1)} = \sqrt{(1 - Sensibilità)^2 + (1 - Specificità)^2};$$

- il criterio di Youden o della massima distanza dalla bisettrice, calcolato come:

$$D_{line} = Sensibilità + Specificità - 1.$$

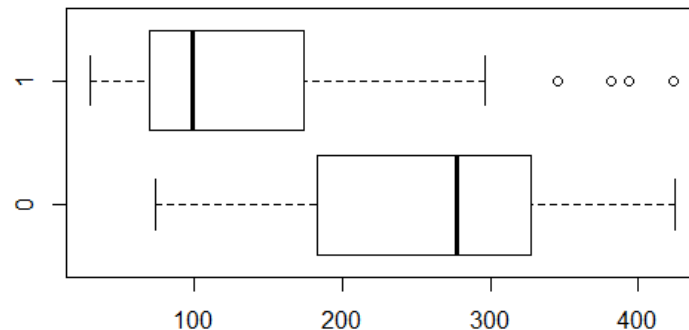
3.3.2 Applicazione della curva ROC al dataset

Considerando nuovamente la variabile follow-up come *golden test*, si vuole provare a individuare un qualche altro criterio (considerando le variabili quantitative rilevate) che permetta di ottenere una classificazione dei pazienti quanto più vicina possibile a quella che si ha impiegando la variabile follow-up.

Le variabili quantitative contenute nel dataset sono età e tempo di mantenimento della cannula, ma, alla luce della non significatività della relazione tra età e follow-up, solamente la variabile tempo potrebbe essere considerata come variabile criterio per

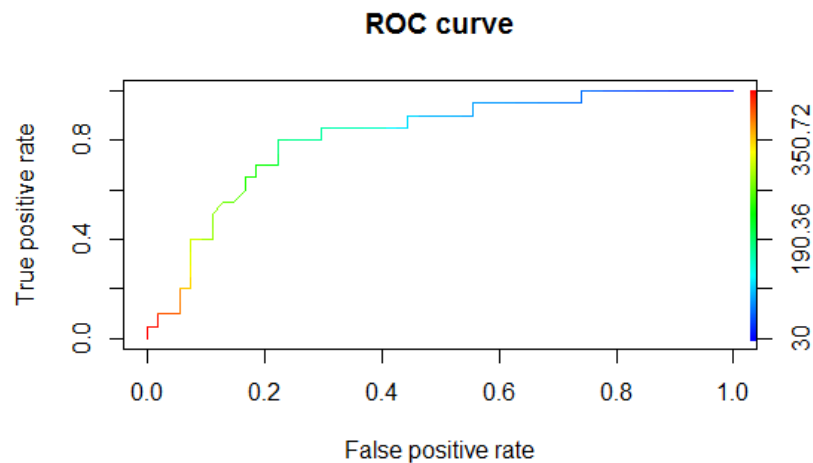
individuare un opportuno valore discriminante tra i due gruppi. Tale variabile è risultata avere una relazione significativa con l'esito della rimozione della cannula a 48 ore; ciò significa che la variabile tempo si distribuisce diversamente nei due gruppi (nel *Capitolo 2* è stata rifiutata l'ipotesi nulla eseguendo il test non parametrico di Mann-Whitney poiché l'ipotesi di normalità veniva accettata in uno solo dei due gruppi). In *Figura 8* sono stati riportati i due boxplot, dai quali si nota una sovrapposizione tra le due distribuzioni; ciò implica che, nel momento in cui si andrà a determinare un valore soglia, si avranno dei pazienti classificati in maniera sbagliata (i Falsi Positivi e i Falsi Negativi). Si osserva anche che il gruppo che non supera il test (0) presenta un range più ampio, mentre l'altro gruppo (1) presenta un range più ristretto, con una distribuzione concentrata verso valori più bassi. La variabilità appare maggiore nel gruppo dei pazienti che non hanno superato il test, ma nell'altro gruppo risulta influenzata da alcuni outlier localizzati verso valori più elevati del dominio.

Figura 8: Boxplot per tempo di mantenimento della cannula distintamente per il gruppo di pazienti che non hanno superato la prova (0) e per il gruppo di pazienti che l'hanno superata (1).



L'area sotto la curva ROC (*Figura 9*) è pari a 0.812 e tale valore indica che il test è moderatamente accurato. L'intervallo di confidenza al 95% per la stessa misura è (0.705, 0.919), per cui l'ipotesi nulla $H_0: AUC=0.5$ viene rifiutata ad un livello di significatività del 5% ($p\text{-value}<0.001$).

Figura 9



Il punto della curva ROC che si trova a massima distanza dalla bisettrice (criterio di Youden) si ha in corrispondenza di un tempo di mantenimento della cannula pari a 178.50 giorni e tale punto, grazie alla regolarità della curva, coincide con quello che ha minima distanza dal punto di ottimalità (0,1).

Utilizzando come valore soglia 178.50 giorni si ottiene la *Tabella 8*.

Tabella 8: Matrice di confusione

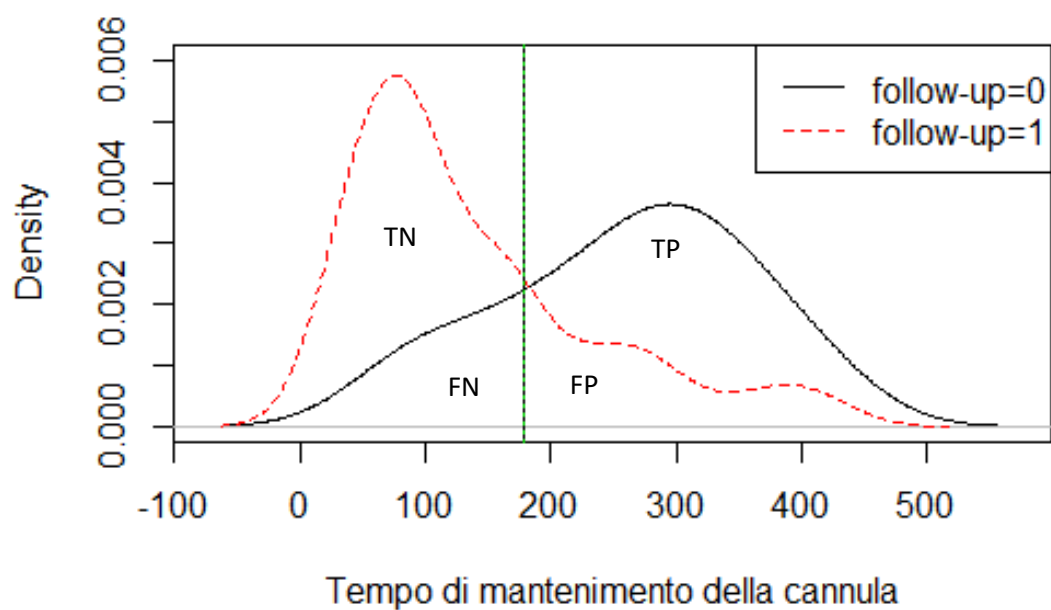
	Follow-up=0	Follow-up=1	Totale
Tempo>178.50	16	12	28
Tempo≤178.50	4	42	46
Totale	20	54	74

La soglia individuata porta ad avere una sensibilità pari all'80% ed una specificità pari al 77.8%, quindi la soluzione proposta risulta essere più sensibile che specifica. Il 20% dei pazienti che non superano il test (follow-up=0) vengono classificati come negativi (FN), mentre il 22.2% dei pazienti che superano il test vengono classificati come positivi (FP).

L'accuratezza, ossia il grado di accordo tra le due misure, è pari al 78.4%.

La *Figura 10* mette in evidenza quanto affermato.

Figura 10: Densità della variabile tempo di mantenimento della cannula distintamente per i due gruppi di pazienti.



CAPITOLO 4

MODELLI DI REGRESSIONE LOGISTICA

4.1 I GLM e il modello logistico

In questo paragrafo si desidera introdurre il modello logistico. Esso viene impiegato nel momento in cui si vuole modellare la media di una variabile risposta binaria, ossia la sua probabilità di successo, in funzione di alcune variabili esplicative (quantitative e/o qualitative).

Il modello di regressione logistica appartiene alla famiglia dei modelli lineari generalizzati (GLM, *Generalized Linear Models*) introdotti da Nelder e Wedderburn (1972). I GLM rappresentano un'estensione del modello di regressione lineare normale per trattare risposte diverse dalla normale (nel caso in esame la variabile risposta ha distribuzione bernoulliana).

La variabile risposta ha distribuzione appartenente ad una famiglia di dispersione esponenziale e, per diverse specificazioni delle funzioni in essa contenute, si trovano, tra i casi particolari più interessanti, le distribuzioni normale, gamma, binomiale e Poisson.

I modelli lineari generalizzati sono caratterizzati dal fatto che la relazione tra variabili esplicative e variabile risposta si può esprimere nella forma

$$g(E\{Y\}) = x^T \beta,$$

per un'opportuna scelta della funzione di legame $g(\cdot)$, derivabile con continuità e invertibile e che collega la media della risposta al predittore lineare. I valori delle variabili x_j sono predeterminati oppure si opera condizionatamente ai valori assunti dalle variabili (Azzalini e Scarpa, 2004).

Per quanto concerne gli aspetti inferenziali, esiste una teoria dell'inferenza sull'argomento ben articolata e che fa riferimento alla verosimiglianza. In generale, la stima di massima verosimiglianza di un GLM si ottiene mediante procedure iterative.

Nel caso della regressione logistica, Y può assumere i valori $\{0,1\}$ ed è una variabile di Bernoulli. La probabilità che Y_i assuma valore pari a 1 è π_i , mentre la probabilità che si verifichi l'evento complementare è pari a $1 - \pi_i$. La media coincide con la probabilità di successo e rappresenta la quantità che si vuole modellare attraverso la regressione.

Si pone:

$$g(\pi_i) = \beta^T x_i = \beta_1 x_{i1} + \dots + \beta_p x_{ip},$$

con $g(\cdot): [0,1] \rightarrow \mathbb{R}$. La regressione logistica è basata sulla scelta della funzione di legame logit (o logistica):

$$\text{logit}(\pi_i) = \log \frac{\pi_i}{1 - \pi_i} = \beta^T x_i;$$

mentre l'inversa è data da:

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

Nel complesso si ha:

$$g(\pi_i) = \log \frac{\pi_i}{1 - \pi_i} = \beta_1 x_{i1} + \dots + \beta_p x_{ip}, \quad \text{per } i = 1, \dots, n.$$

La funzione di legame logit comporta notevoli vantaggi:

- è interpretabile in termini di log-rapporto delle probabilità di successo e insuccesso (il generico coefficiente β_r esprime l'effetto sul log-rapporto della quota di un incremento unitario di x_{ir} , fermo restando il valore delle ulteriori esplicative nel modello);
- permette semplificazioni nelle quantità di verosimiglianza, agevolando le analisi statistiche;
- per gli studi di tipo epidemiologico, essa permette di trattare dati raccolti retrospettivamente.

In sintesi, nella regressione logistica si assume:

- 1) $Y_1, Y_2 \dots Y_n$ variabili indipendenti, con $Y_i \sim \text{Bernoulli}(\pi_i)$;
- 2) $X = (x_1, \dots, x_n)^T$, di dimensioni $(n \times p)$, non stocastica e con rango pieno;
- 3) legame $g(\pi_i) = \text{logit}(\pi_i) = \beta^T x_i$;
- 4) eteroschedasticità:

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

4.2 Il problema di separazione

Nell'applicare il modello logistico al dataset, in particolare considerando la variabile follow-up (dicotomica) come variabile risposta, emerge che non si riesce a costruire un modello soddisfacente (si ottengono infatti stime e *standard errors* particolarmente elevati). Questo risultato è dovuto ad un problema di separazione delle variabili (Heinze e Schemper, 2002).

Il fenomeno di separazione si presenta principalmente con campioni di piccole dimensioni ed è caratterizzato dalla monotonia della funzione di verosimiglianza. Quando si manifesta questo problema, generalmente il processo di stima del modello logistico porta ad ottenere una stima divergente per almeno un parametro (tuttavia si potrebbe anche manifestare se le stime dei parametri inclusi nel modello sono basse in valore assoluto). Si tratta di un problema di inesistenza della stima di massima verosimiglianza sotto speciali condizioni del campione e non è affatto trascurabile. Inoltre la probabilità del verificarsi di un evento di separazione dipende non solo dalla dimensione del campione, bensì anche dal numero dei fattori di rischio dicotomici, dalla grandezza dei rapporti tra quote (*odds ratio*) ad essi associati e dal grado di equilibrio nella loro distribuzione (Heinze e Schemper, 2002).

In generale, esistono tre tipologie di configurazione dei dati esaustive e mutuamente esclusive: la separazione completa, la separazione quasi-completa e la sovrapposizione. Le prime due producono stime non uniche ed infinite, mentre la terza produce sempre stime uniche e finite.

A titolo esemplificativo, la *Figura 11* mostra le possibili configurazioni delle osservazioni nel caso di due variabili, x_1 e x_2 , quantitative e due gruppi (uno contraddistinto dai pallini e uno dalle crocette). In base alla regola di allocazione, vengono definite le regioni R_1 e R_2 .

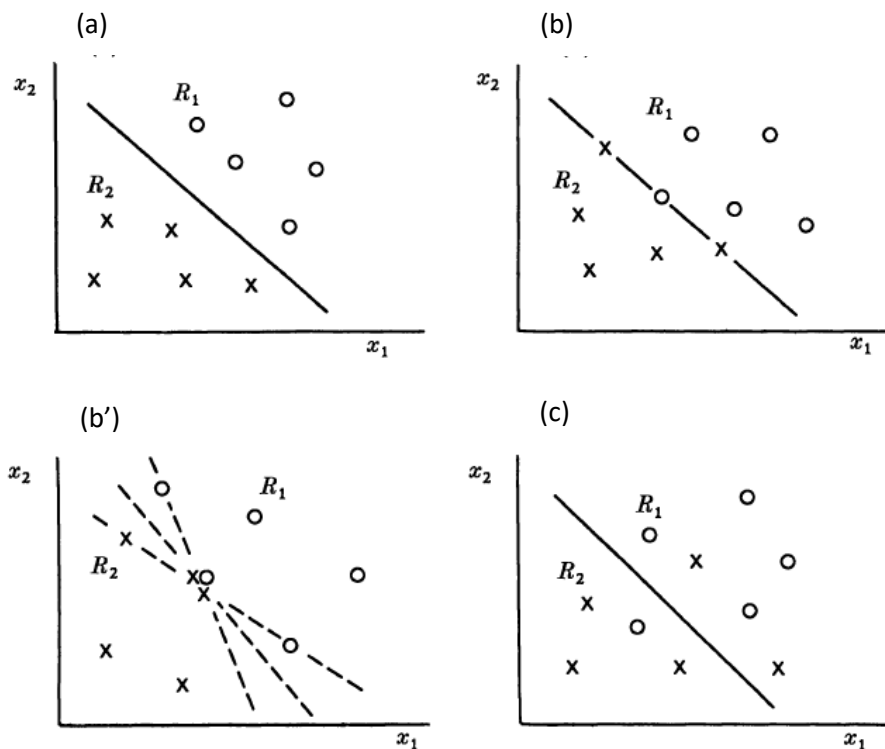
Si ha separazione completa (*Figura 11a*) nel momento in cui un fattore di rischio o una combinazione lineare di più fattori di rischio alloca correttamente tutte le osservazioni al loro gruppo di appartenenza. In altre parole, considerando una tabella di

contingenza 2x2 tra la variabile risposta e un fattore di rischio (entrambi dicotomizzati) emerge che tutte le frequenze si trovano lungo la diagonale della tabella.

Invece, la separazione quasi completa (*Figura 11b e 11b'*) avviene quando quasi tutte le osservazioni vengono classificate in maniera corretta.

Il caso di sovrapposizione (*Figura 11c*), infine, vede la presenza di osservazioni di entrambi i gruppi in entrambe le regioni R_1 e R_2 .

Figura 11: Possibili configurazioni. (a) separazione completa; (b) e (b') separazione quasi completa; (c) sovrapposizione.



Quando la separazione è verificata, i test di Wald e gli intervalli di confidenza sono disponibili ma poco affidabili (Heinze e Schemper, 2002).

Le soluzioni che verranno trattate allo scopo di risolvere il problema di separazione sono la procedura sviluppata da Firth (1993) e una procedura più recente sviluppata da Sartori *et al.* (2017).

Nei prossimi paragrafi vengono descritte teoricamente le metodologie statistiche proposte ed applicate al dataset studiato.

4.3.1 Approcci standard per la riduzione della distorsione (*bias*) della stima di massima verosimiglianza

Una delle proprietà più importanti dello stimatore di massima verosimiglianza, ottenuto come soluzione della funzione di punteggio (*score function*), è la non distorsione asintotica; tuttavia nei campioni finiti lo stimatore è in generale distorto, con distorsione di ordine $O(n^{-1})$.

In un modello statistico parametrico regolare con parametro θ p -dimensionale, ossia $\theta = (\theta_1, \dots, \theta_p)$, la distorsione asintotica dello stimatore di massima verosimiglianza di θ può essere sviluppata come segue:

$$b(\theta) = \frac{b_1(\theta)}{n} + \frac{b_2(\theta)}{n^2} + \dots,$$

dove n è solitamente dato dal numero di osservazioni, ma può essere anche qualche altra misura del tasso al quale si accumula l'informazione (Firth, 1993), mentre $b_1(\theta), b_2(\theta), \dots$ è un'appropriata sequenza di funzioni, ognuna delle quali è di ordine $O(1)$ (Kosmidis, 2014).

In letteratura sono stati ampiamente studiati due approcci standard. Il primo è il metodo *Jackknife* (Quenouille, 1949, 1956), intensamente computazionale; è molto generale e non richiede il calcolo di $\frac{b_1(\theta)}{n}$ per la sua implementazione. Tuttavia, la mancata richiesta di calcoli teorici compensa con una perdita in termini di precisione. L'idea su cui si basa è la seguente: dato un campione casuale semplice con $n > 1$ osservazioni, si suppone di avere a disposizione tutte le informazioni tranne la j -esima. Con l'informazione che si ha, si calcola $\hat{\theta}_{(-j)}$. Quenouille ha mostrato che lo stimatore $\tilde{\theta} = n\hat{\theta} - (n-1)\bar{\theta}$, dove $\bar{\theta}$ rappresenta la media degli n possibili stimatori $\hat{\theta}_{(-1)}, \hat{\theta}_{(-2)}, \dots, \hat{\theta}_{(-n)}$ ottenuti rimuovendo dal campione l'osservazione j -esima per $j = 1, \dots, n$, ha distorsione pari a $-\frac{b_2(\theta)}{n^2} + O(n^{-3})$, cioè di ordine asintotico più basso rispetto allo stimatore $\hat{\theta}$ (Kosmidis, 2014).

Il secondo approccio sostituisce semplicemente la stima di massima verosimiglianza $\hat{\theta}$ a θ in $b_1(\theta)$. La stima corretta dalla distorsione è quindi calcolata come:

$$\hat{\theta}_{BC} = \hat{\theta} - \frac{b_1(\hat{\theta})}{n}.$$

Questo stimatore ha distorsione di ordine $O(n^{-2})$ ed è in generale efficiente.

Entrambi questi metodi funzionano rimuovendo il termine $\frac{b_1(\theta)}{n}$ dalla distorsione asintotica. Un'altra comune caratteristica di questi due approcci consiste nel fatto che sono “correttivi”, piuttosto che “preventivi”. Ciò significa che la stima di massima verosimiglianza $\hat{\theta}$ viene prima calcolata e, solo in un secondo momento, corretta.

Un requisito pratico per l'applicazione di qualsiasi metodo ad un campione finito è l'esistenza di $\hat{\theta}$ per il campione. Nella pratica, in particolare per campioni di piccola o media dimensione, può accadere che la stima diverga; come visto i modelli logistici per risposte binarie sono inclini a questo comportamento.

Essendo necessario un approccio per la riduzione della distorsione che non dipenda dal fatto che la stima di massima verosimiglianza sia finita, Firth (1993) ha proposto un metodo che si differenzia dagli altri due per l'applicazione di una correzione sistematica per il meccanismo che produce la stima di massima verosimiglianza, ossia l'equazione di verosimiglianza basata sulla funzione di punteggio (*score function*), piuttosto che agire direttamente sulla stima.

4.3.2 Il metodo di Firth

Nei problemi regolari di stima, lo stimatore di massima verosimiglianza $\hat{\theta}$ si ottiene come soluzione dell'equazione basata sulla funzione di punteggio:

$$\ell_*(\theta) = \frac{\partial \ell(\theta)}{\partial \theta} = 0,$$

dove $\ell(\theta) = \log L(\theta)$ è la funzione di log-verosimiglianza.

Si consideri inizialmente (Firth, 1993) un modello della famiglia esponenziale con $\ell(\theta) = t\theta - K(\theta)$, con parametro θ scalare (cioè $p=1$). Derivando la funzione di log-verosimiglianza si ottiene:

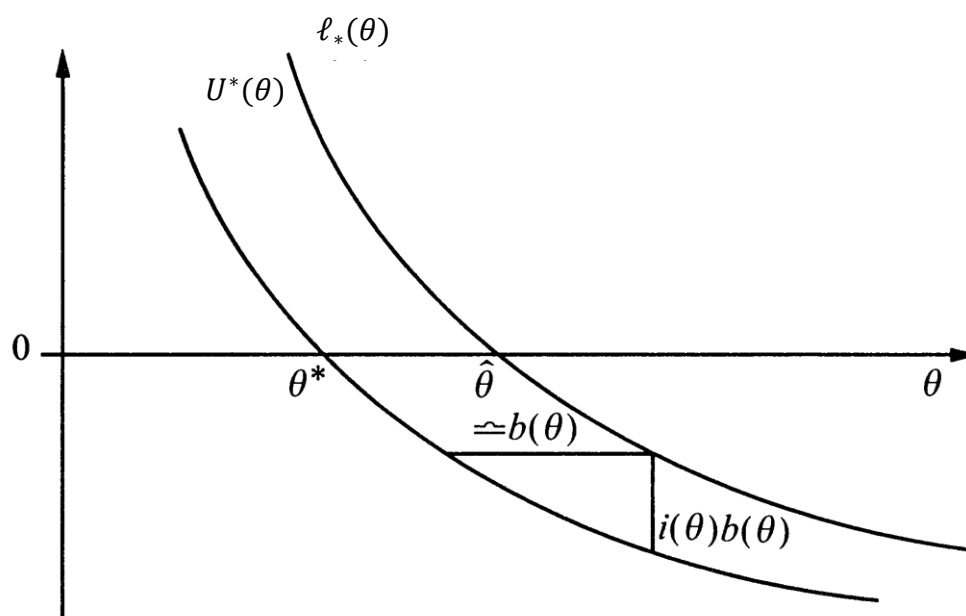
$$\ell_*(\theta) = t - K'(\theta),$$

dove si nota che la statistica sufficiente t interessa solo la posizione di $\ell_*(\theta)$, non la sua forma.

La distorsione di $\hat{\theta}$ nasce dalla combinazione di due fattori: uno dovuto alla non distorsione della funzione di punteggio, per cui $E\{\ell_*(\theta)\} = 0$ per il vero valore del parametro, mentre l'altro è dovuto alla curvatura della funzione di punteggio ($\ell_{**}(\theta) \neq 0$).

Se $\ell_*(\theta)$ è lineare in θ allora $E(\hat{\theta}) = \theta$; ma la curvatura combinata con la non distorsione della funzione di punteggio (Figura 12) possono portare alla distorsione di $\hat{\theta}$ (nella direzione positiva nel caso illustrato).

Figura 12: Modifica della funzione di punteggio non distorta



L'idea fondamentale di Firth è che la distorsione di $\hat{\theta}$ possa essere ridotta introducendo una piccola distorsione nella funzione di punteggio.

Se $\hat{\theta}$ è soggetto a una distorsione $b(\theta)$ positiva (cioè se $E(\hat{\theta}) > \theta$), la funzione di punteggio è spostata verso il basso per ogni punto di θ di un ammontare pari a $i(\theta)b(\theta)$, dove $-i(\theta) = E\{\ell_{**}(\theta)\}$ è il gradiente locale; si giunge alla determinazione di una funzione di punteggio modificata:

$$U^*(\theta) = \ell_*(\theta) - i(\theta)b(\theta)$$

e quindi ad una stima modificata θ^* , soluzione di $U^*(\theta) = 0$.

Nel caso si disponga di un vettore di parametri, essa può essere interpretata come un vettore di equazioni, mentre $i(\theta)$ rappresenta la matrice di informazione di Fisher.

Per generalizzare il caso appena analizzato a diversi modelli, è utile fare riferimento alla notazione introdotta da McCullagh (1987) per quanto riguarda le derivate della log-verosimiglianza e i loro momenti.

Le derivate vengono descritte come

$$U_r(\theta; Y_i) = \frac{\partial \ell(\theta; Y_i)}{\partial \theta_r}, \quad U_{rs}(\theta; Y_i) = \frac{\partial^2 \ell(\theta; Y_i)}{\partial \theta_r \partial \theta_s},$$

e così via, dove $\theta = (\theta_1, \dots, \theta_p)$ è il vettore dei parametri.

I momenti sono definiti come

$$\kappa_{r,s} = n^{-1}E\{U_r U_s\}, \quad \kappa_{r,s,t} = n^{-1}E\{U_r U_s U_t\}, \quad \kappa_{r,st} = n^{-1}E\{U_r U_{st}\}$$

e così via.

Sono note inoltre le seguenti relazioni:

$$\kappa_{rs} + \kappa_{r,s} = 0, \quad \kappa_{rst} + \kappa_{r,st} + \kappa_{t,rs} + \kappa_{r,s,t} = 0.$$

Si consideri ora una modifica abbastanza generale della funzione di punteggio della forma:

$$U_r^*(\theta) = U_r(\theta) + A_r(\theta),$$

dove A_r può dipendere dai dati ed è $O_p(1)$ per $n \rightarrow \infty$.

Si supponga che $\hat{\theta}$ e θ^* soddisfino, rispettivamente, $U(\hat{\theta}) = 0$ e $U^*(\theta^*) = 0$ e si ipotizzi $\hat{\gamma} = n^{\frac{1}{2}}(\theta^* - \theta)$. Allora, se ci si basa su un'estensione di $U_r^*(\theta^*)$, la distorsione di θ^* è:

$$E\left(n^{-1/2}\hat{\gamma}^r\right) = n^{-1}\kappa^{r,s}\left\{-\frac{\kappa^{t,u}(\kappa_{s,t,u} + \kappa_{s,tu})}{2} + \alpha_s\right\} + O\left(n^{-3/2}\right),$$

dove $\kappa^{r,s}$ rappresenta l'inversa della matrice dell'informazione di Fisher $\kappa_{r,s}$, mentre α_s è il valore atteso nullo di A_s .

La distorsione del primo ordine di $\hat{\theta}$ è data dal termine:

$$-\frac{n^{-1}\kappa^{r,s}\kappa^{t,u}(\kappa_{s,t,u} + \kappa_{s,tu})}{2} = n^{-1}b_1^r(\theta).$$

A_r rimuove il termine del primo ordine se soddisfa

$$\kappa^{r,s}\alpha_s = -b_1^r + O\left(n^{-1/2}\right).$$

Si ottiene

$$\alpha_r = -\kappa_{r,s}b_1^s + O\left(n^{-1/2}\right).$$

In notazione matriciale il vettore $A = [\alpha_r]$ dovrebbe avere una forma tale per cui

$$E(A) = -\frac{i(\theta)b_1(\theta)}{n} + O\left(n^{-1/2}\right).$$

Ovvi candidati per la scelta di A sono:

$$A^{(E)} = -\frac{i(\theta)b_1(\theta)}{n} \text{ e } A^{(O)} = -\frac{j(\theta)b_1(\theta)}{n},$$

utilizzando, rispettivamente, l'informazione attesa $i(\theta)$ e l'informazione osservata $j(\theta)$. Nel caso di famiglia di dispersione esponenziale con la parametrizzazione canonica l'informazione osservata non coinvolge i dati, per cui $A^{(O)}$ e $A^{(E)}$ coincidono. In generale, entrambe queste modificazioni portano alla rimozione del termine di distorsione $O(n^{-1})$.

4.3.3 Applicazione del metodo di Firth al dataset

Le funzioni che sono state utilizzate per lo svolgimento delle analisi sono contenute nella libreria *logistf* di R. Essa permette di stimare un modello logistico usando il metodo per la riduzione della distorsione proposto da Firth. Il modello stimato appartiene alla classe *logistf* e fornisce la stima dei parametri, gli *standard errors*, gli intervalli di confidenza per ogni parametro, i *p-values* per la verifica dell'ipotesi di nullità per ogni parametro, il valore della massima log-verosimiglianza, i predittori lineari e il numero di iterazioni che sono state necessarie per la convergenza dell'algoritmo. Gli intervalli di confidenza per i coefficienti del modello possono essere calcolati attraverso la verosimiglianza profilo, seguendo l'algoritmo di Venzon e Moolgavkar (1988).

I test del log-rapporto di verosimiglianza sono basati sulla funzione

$$2 \left(\ell(\hat{\theta}^*) - \ell(\theta) \right),$$

dove $\ell(\theta)$ è la funzione di log-verosimiglianza e $\hat{\theta}^*$ è la stima di massima verosimiglianza penalizzata secondo il metodo di Firth (Heinze e Schemper, 2002).

Sono stati stimati tre modelli per prevedere la variabile risposta follow-up. Si è sempre partiti dal modello completo ed è stata successivamente seguita una procedura all'indietro (*backward elimination*), eliminando di volta in volta la variabile che presentava il valore del test del rapporto di verosimiglianza minore o, equivalentemente, il *p-value* associato al test superiore.

Il primo modello stimato vuole modellare la probabilità di successo dell'evento di decannulazione considerando come modello di partenza quello che include le variabili esplicative che riguardano i test che sono stati eseguiti (non i tre raggruppamenti dei test) e supponendo di non conoscere la variabile esito.

Il modello finale a cui si è giunti include le variabili tosse riflessa, chiusura e pervietà (Tabella 9). Il modello stimato è

$$\log\left(\frac{\hat{\pi}_i}{1 - \hat{\pi}_i}\right) = -14.32 + 4.67 * \text{tosse riflessa}_i + 7.71 * \text{chiusura}_i + 6.61 * \text{pervietà}_i.$$

In particolare si nota che ogni coefficiente angolare ha segno positivo e ciò si traduce nel fatto che l'aver superato almeno uno tra i tre test usati come esplicative porta ad un incremento del log-rapporto della quota e quindi della probabilità di successo della rimozione della cannula.

Tabella 9: Tabella dei coefficienti del primo modello (Firth)

Parametro	Stima	Standard Error	Chisq value	p-value
<i>Intercetta</i>	-14.321084	3.929	31.317	<0.001
<i>Tosse riflessa</i>	4.665412	2.007	6.495	0.011
<i>Chiusura</i>	7.710221	2.078	50.706	<0.001
<i>Pervietà</i>	6.614722	2.255	20.047	<0.001

Il test del log-rapporto di verosimiglianza porta al rifiuto dell'ipotesi di uguaglianza tra il modello corrente e il modello nullo ($W^{\text{oss}}=73.86$, $df=3$, $\alpha^{\text{oss}}<0.001$).

Per valutare la bontà del modello è possibile costruire una tabella di contingenza (o matrice di confusione) tra la variabile follow-up e i valori predetti dal modello, dove il successo (1) è rappresentato da un $\hat{\pi}_i \geq 0.5$ (Tabella 10).

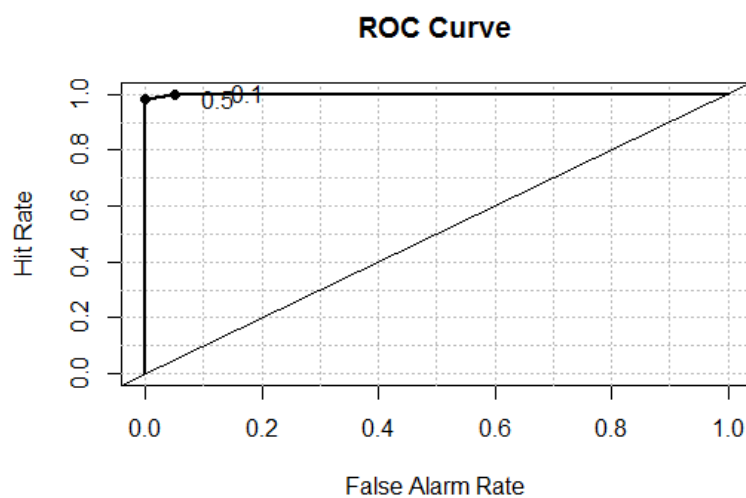
Tabella 10: Matrice di confusione tra la variabile follow-up e i valori predetti dal primo modello

	Valore predetto=0	Valore predetto=1
Follow-up=0	19	1
Follow-up=1	0	54

L'accuratezza è pari al 98.6%, la sensibilità (ossia il rapporto tra coloro che hanno follow-up=0 e registrano un insuccesso secondo il modello e il numero totale di soggetti che hanno follow-up=0) al 95% e la specificità (ossia il rapporto tra coloro che hanno follow-up=1 e registrano un successo secondo il modello e il numero totale di

soggetti che hanno follow-up=1) è la massima possibile (100%). Si tratta di un modello molto buono, in quanto prevede in maniera sbagliata una sola osservazione (in particolare il modello prevede che tale soggetto supererà la decannulazione, mentre il valore registrato dalla variabile follow-up è pari a 0, si tratta cioè di un falso negativo). La *Figura 13* riporta la curva ROC per la valutazione dell'adeguatezza del modello. Con una stima non parametrica dell'area sotto la curva (AUC) pari a 0.9995 (I.C. al 95%: 0.9983-1), indica che il modello prevede in modo altamente accurato le osservazioni.

Figura 13



Il secondo modello che si è stimato segue la stessa logica del primo per quanto riguarda le variabili esplicative incluse nel modello iniziale, ad eccezione della variabile esito, la quale in questo caso si assume nota. Vengono perciò nuovamente escluse le variabili ottenute come combinazione dei vari test (i gruppi).

Il modello finale stimato è

$$\log\left(\frac{\hat{\pi}_i}{1 - \hat{\pi}_i}\right) = -10 + 6.64 * pervietà_i + 8.06 * esito_i.$$

Tabella 11: Tabella dei coefficienti del secondo modello (Firth)

Parametro	Stima	Standard Error	Chisq value	p-value
<i>Intercetta</i>	-10.003179	2.703	34.812	<0.001
<i>Pervietà</i>	6.636338	2.258	20.176	<0.001
<i>Esito</i>	8.057737	2.066	62.50	<0.001

Il test del log-rapporto di verosimiglianza porta a preferire il modello corrente rispetto a quello nullo ($W^{oss}=77.29$, $df=2$, $\alpha^{oss}=0$).

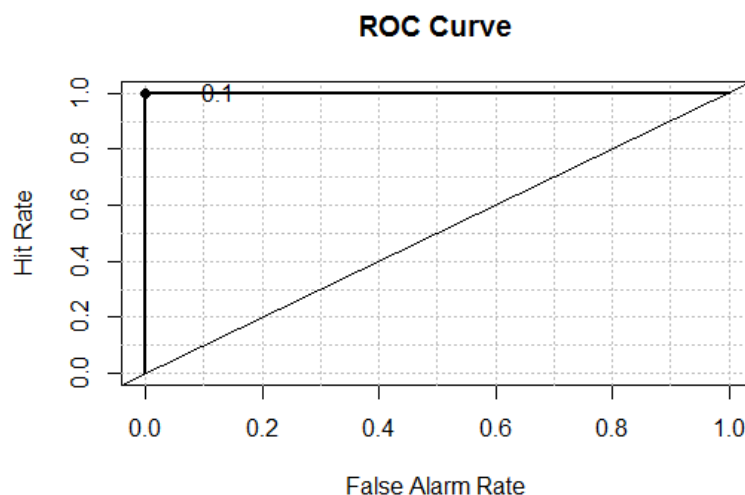
La *Tabella 12* mette in evidenza il fatto che tutti i pazienti vengono previsti in maniera corretta, per cui i valori di accuratezza, sensibilità e specificità sono massimi (100%).

Tabella 12: Matrice di confusione tra la variabile follow-up e i valori predetti dal secondo modello

	Valore predetto=0	Valore predetto=1
Follow-up=0	20	0
Follow-up=1	0	54

Anche la curva ROC (*Figura 14*) fa notare che si tratta di un modello perfetto (AUC=1).

Figura 14



Il terzo modello stimato vuole invece modellare la probabilità di riuscita della decannulazione dopo 48 ore sulla base delle esplicative ottenute come raggruppamento di più test (gruppo pervietà, gruppo disfagia, gruppo clinico), partendo sempre dal modello completo.

Il modello finale stimato è

$$\log\left(\frac{\hat{\pi}_i}{1 - \hat{\pi}_i}\right) = -3.66 + 7.26 * \text{gruppo pervietà}_i.$$

Tabella 13: Tabella dei coefficienti del terzo modello (Firth)

Parametro	Stima	Standard Error	Chisq value	p-value
<i>Intercetta</i>	-3.663562	1.469	23.04961	<0.001
<i>Gruppo pervietà</i>	7.256297	1.690	Inf	0

Il test del log-rapporto di verosimiglianza mostra evidenza a favore del modello corrente rispetto a quello nullo ($W^{oss}=71.10$, $df=1$, $\alpha^{oss}=0$).

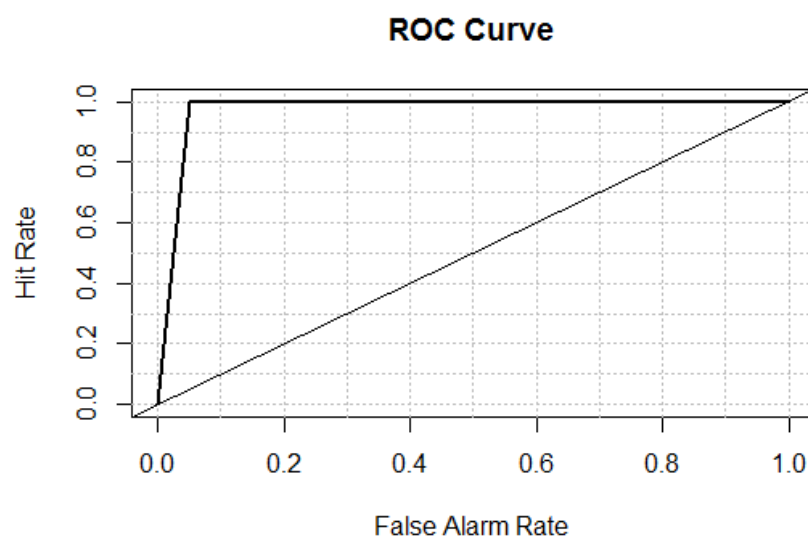
Dalla *Tabella 14* si deduce che anche questo modello è molto buono, in quanto prevede in maniera errata un solo paziente. I valori di accuratezza, sensibilità e specificità sono, rispettivamente, 98.6%, 95% e 100%.

Tabella 14: Matrice di confusione tra la variabile follow-up e i valori predetti dal terzo modello

	Valore predetto=0	Valore predetto=1
Follow-up=0	19	1
Follow-up=1	0	54

La curva ROC (*Figura 15*) presenta una stima non parametrica dell'AUC pari a 0.975 (I.C. 95%: 0.926-1) e graficamente si può notare la distanza dalla bisettrice. Il modello prevede in modo altamente accurato le osservazioni.

Figura 15



Tutti i modelli proposti sono molto buoni, anche se tra questi spicca il secondo modello in termini di accuratezza nella previsione. L'AIC, il criterio di informazione di Akaike per la valutazione della perdita attesa di efficacia predittiva del modello, risulta essere minimo proprio nel secondo modello; si ha infatti: -67.86 per il primo modello, -73.29 per il secondo modello e -69.10 per il terzo modello. Analogamente, confrontando i valori del test del log-rapporto di verosimiglianza per il confronto tra modello corrente e modello nullo si giunge alla medesima conclusione, in quanto il secondo modello è quello che si allontana maggiormente dal modello con la sola intercetta.

Nel caso in cui non fosse nota la variabile esito, i modelli disponibili sarebbero il primo e il terzo; tra i due quello che include solo la variabile gruppo pervietà presenta AIC minimo. Anche per quanto riguarda il numero dei parametri ($p=4$ nel primo e $p=2$ nel terzo), essendo il terzo modello più parsimonioso, risulta preferibile al primo. Invece il test del log-rapporto di verosimiglianza per il confronto tra il modello corrente e il modello nullo porta alla conclusione opposta (il primo modello ha un valore del test più alto rispetto al terzo). Tale differenza è dovuta al fatto che l'AIC penalizza la log-verosimiglianza tramite un termine che dipende dal numero di parametri inclusi nel modello. Anche il confronto tramite AUC porta a preferire il primo modello.

Si nota infine che in tutti i modelli è presente la variabile pervietà e da ciò si deduce un suo contributo significativo nel prevedere la riuscita della decannulazione.

4.3.4 La riduzione della distorsione in mediana della stima di massima verosimiglianza

Il limite principale nel metodo di Firth consiste nel fatto che la riduzione della distorsione è strettamente legata alla parametrizzazione scelta. Viene quindi proposta una modifica in mediana dell'equazione di stima, la cui soluzione rispetta l'equivarianza sotto riparametrizzazioni monotone (Sartori *et al.*, 2017).

Come accadeva con il metodo di Firth, anche questa procedura non richiede che la stima di massima verosimiglianza sia finita ed è efficace nel prevenire stime infinite. La modifica si ottiene considerando la mediana, al posto della media, come un indice di centratura della funzione di punteggio e definendo una nuova funzione di stima tramite sottrazione dalla funzione di punteggio della sua mediana approssimativa.

La centratura mediana della funzione di punteggio implica la centratura mediana del corrispondente stimatore, purché la nuova equazione di stima abbia soluzione unica. Allora lo stimatore è approssimativamente non distorto in mediana.

Nella maggior parte dei casi le approssimazioni per le stime non distorte in mediana sono basate su verosimiglianze asintotiche di ordine elevato.

Il caso che verrà presentato nel seguito è quello più semplice, dove θ è un parametro scalare (per una trattazione più approfondita, con riferimento anche al caso di parametro vettoriale, si veda Sartori *et al.*, 2017).

Si consideri per le osservazioni y un modello regolare con funzione di densità di probabilità $p_y(y; \theta)$, $\theta \in \Theta \in \mathbb{R}$. Sia $\ell(\theta)$ la corrispondente funzione di log-verosimiglianza e $\ell_*(\theta) = \frac{\partial \ell(\theta)}{\partial \theta}$ la funzione di punteggio. Lo stimatore di massima verosimiglianza $\hat{\theta}$ si ottiene come soluzione dell'equazione $\ell_*(\theta) = 0$.

Si assume che l'informazione di Fisher $i(\theta)$ e i cumulanti del terzo ordine di $\ell_*(\theta)$ siano finiti e di ordine $O(n)$, dove n rappresenta anche in questo caso la dimensione del campione o un indice dell'informazione contenuta nei dati.

Utilizzando l'espansione di Cornish-Fisher (Pace e Salvani, 1997) si ottiene lo sviluppo asintotico per la mediana $M(\cdot)$ della funzione di punteggio:

$$M\{\ell_*(\theta)\} = -\frac{\nu_{\theta,\theta,\theta}}{\{6i(\theta)\}} + O(n^{-1}),$$

dove $\nu_{\theta,\theta,\theta} = \nu_{\theta,\theta,\theta}(\theta) = E(U(\theta)^3)$.

La funzione di punteggio modificata viene definita come segue:

$$\tilde{U}(\theta) = \ell_*(\theta) + \frac{\nu_{\theta,\theta,\theta}}{\{6i(\theta)\}},$$

dove il termine della modifica $\nu_{\theta,\theta,\theta}$ è di ordine $O(1)$. Con $\tilde{\theta}$ viene indicata la soluzione dell'equazione $\tilde{U}(\theta) = 0$.

Per $\tilde{U}(\theta)$ si ha $M\{\tilde{U}(\theta)\} = O(n^{-1})$ e risulta non distorta in mediana al terzo ordine; infatti:

$$P_{\theta}(\tilde{U}(\theta) \leq 0) = \frac{1}{2} + O\left(n^{-\frac{3}{2}}\right).$$

Se $\tilde{\theta}$ è l'unica soluzione di $\tilde{U}(\theta) = 0$, gli eventi $\tilde{U}(\theta) \leq 0$ e $\tilde{\theta} \leq \theta$ sono equivalenti, cosicché $\tilde{\theta}$ è non distorto in mediana al terzo ordine:

$$P_{\theta}(\tilde{\theta} \leq \theta) = \frac{1}{2} + O\left(n^{-\frac{3}{2}}\right).$$

$\tilde{\theta}$ ha distribuzione asintotica $N(\theta, i(\theta)^{-1})$ e quindi gli intervalli di confidenza di Wald possono essere calcolati e differiscono solo per la posizione rispetto a quelli calcolati per $\hat{\theta}$.

Si consideri ora una riparametrizzazione $\omega(\theta)$ con inversa $\theta(\omega)$. Le quantità precedentemente definite assumono la forma:

$$v_{\omega, \omega, \omega}^{\Omega} = v_{\theta, \theta, \theta} \{\theta(\omega)\} \{\theta'(\omega)\}^3, \quad i^{\Omega}(\omega) = i\{\theta(\omega)\} \{\theta'(\omega)\}^2,$$

dove $\theta'(\omega) = \frac{\partial \theta(\omega)}{\partial \omega}$.

Sempre secondo la nuova parametrizzazione, $\tilde{U}(\theta)$ diventa $\tilde{U}\{\theta(\omega)\}\theta'(\omega)$ e $\tilde{\omega} = \omega(\tilde{\theta})$ è anch'essa non distorta in mediana al terzo ordine.

La modifica della funzione di punteggio proposta da Firth per ridurre la distorsione in media nel caso scalare è:

$$U^*(\theta) = \ell_*(\theta) + \frac{v_{\theta, \theta, \theta} + v_{\theta, \theta \theta}}{\{2i(\theta)\}},$$

con $v_{\theta, \theta \theta} = E_{\theta}\{\ell_*(\theta)\ell_{**}(\theta)\}$, con $\ell_{**}(\theta) = \frac{\partial^2 \ell(\theta)}{\partial \theta^2}$. Kosmidis & Firth (2010) mostrano che $U^*(\theta)$ non si comporta in modo tensoriale rispetto a riparametrizzazioni.

Se Y ha distribuzione all'interno della famiglia esponenziale con parametro canonico θ si ha:

$$p_Y(y; \theta) = \exp\{\theta t(y) - K(\theta)\} h(y).$$

La funzione di punteggio modificata in mediana ha la forma

$$\tilde{U}(\theta) = \ell_*(\theta) + \frac{K_{\theta\theta\theta}}{\{6K_{\theta\theta}\}},$$

con $K_{\theta\theta\theta} = \frac{\partial^3 K(\theta)}{\partial \theta^3}$ e, in maniera analoga, $K_{\theta\theta} = \frac{\partial^2 K(\theta)}{\partial \theta^2} = i(\theta)$. In questa parametrizzazione, $\tilde{U}(\theta)$ può essere vista come la funzione di punteggio della log-verosimiglianza penalizzata $\tilde{\ell}(\theta) = \ell(\theta) + \frac{\{\log i(\theta)\}}{6}$; ciò implica che l'effetto della modifica in mediana porta a penalizzare la verosimiglianza per un fattore pari a $|i(\theta)|^{\frac{1}{6}}$.

4.3.5 Applicazione del nuovo metodo al dataset

La libreria di R che è stata utilizzata per la stima dei modelli si chiama *mbrglm* e consente di applicare la riduzione della distorsione in mediana nel caso di variabile risposta binomiale, presentando così una soluzione al problema di separazione.

In questo paragrafo i tre modelli precedentemente stimati con il metodo di Firth vengono ristimati con la nuova procedura e presentati nel medesimo ordine con cui sono stati descritti nel paragrafo 4.3.3 per agevolare il successivo confronto tra le due metodologie.

Anche per la stima di questi modelli è stata seguita una procedura di selezione delle variabili all'indietro, partendo perciò dal modello completo. Vengono eliminate una alla volta le variabili che non risultano significative seguendo i valori (minimi) del test di Wald o, equivalentemente, i valori (massimi) del *p-value* associato al test.

La prima proposta consiste nel modellare la probabilità di successo della rimozione della cannula (follow-up) considerando come modello completo quello che include tutte le variabili dicotomiche (ad eccezione di esito e dei gruppi) e quantitative rilevate.

Il modello finale stimato è

$$\log\left(\frac{\hat{\pi}_i}{1 - \hat{\pi}_i}\right) = -15.77 + 5.31 * \text{tosse riflessa}_i + 8.50 * \text{chiusura}_i + 7.38 * \text{pervietà}_i.$$

Tabella 15: Tabella dei coefficienti del primo modello (nuova procedura)

Parametro	Stima	Standard Error	Z-value	p-value
<i>Intercetta</i>	-15.767	4.957	-3.181	0.001
<i>Tosse riflessa</i>	5.305	2.507	2.116	0.034
<i>Chiusura</i>	8.502	2.580	3.296	<0.001
<i>Pervietà</i>	7.381	2.715	2.719	0.007

Dal confronto tra il modello corrente e il modello saturo emerge che i due modelli hanno lo stesso contenuto informativo (Devianza residua=4.9374, *df*=70), preferendo così quello corrente. Quest'ultimo risulta preferibile anche nei confronti del modello nullo ($TRV^{OSS}=81.4251$, *df*=3, $\alpha^{OSS}=0$).

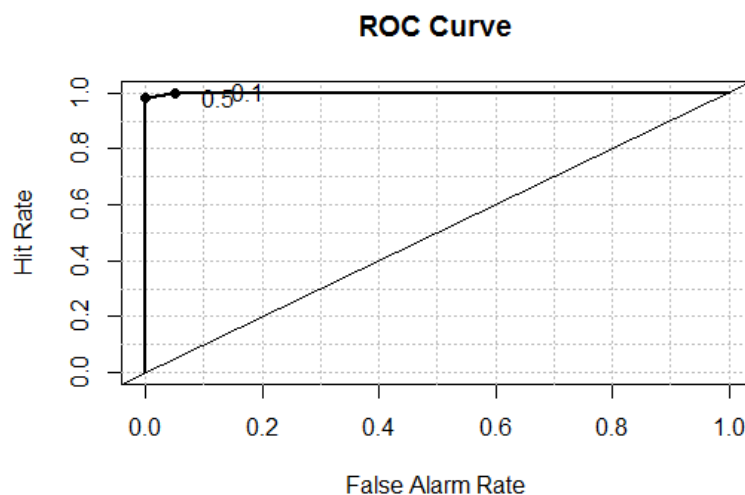
La matrice di confusione (*Tabella 16*) fa notare che solo per un paziente non si riesce a prevedere correttamente l'esito della rimozione della cannula. I valori di accuratezza, sensibilità e specificità sono pari, rispettivamente, a 98.6%, 95% e 100%, evidenziando così la bontà del modello.

Tabella 16: Matrice di confusione tra la variabile follow-up e i valori predetti dal primo modello

	Valore predetto=0	Valore predetto=1
Follow-up=0	19	1
Follow-up=1	0	54

La curva ROC (*Figura 16*) è distante dalla bisettrice; la stima non parametrica dell'area ad essa sottesa è 0.9995 (C.I. 95%: 0.9983-1), indicando che il modello prevede le osservazioni in modo altamente accurato.

Figura 16



Il secondo modello stimato, a differenza del primo, considera nota la variabile esito, in maniera tale da poterla includere nel modello.

Il modello finale stimato è

$$\log\left(\frac{\hat{\pi}_i}{1 - \hat{\pi}_i}\right) = -10.62 + 7.40 * pervietà_i + 8.84 * esito_i.$$

Tabella 17: Tabella dei coefficienti del secondo modello (nuova procedura)

Parametro	Stima	Standard Error	Z-value	p-value
<i>Intercetta</i>	-10.618	3.123	-3.400	<0.001
<i>Pervietà</i>	7.395	2.796	2.645	0.008
<i>Esito</i>	8.837	2.656	3.327	<0.001

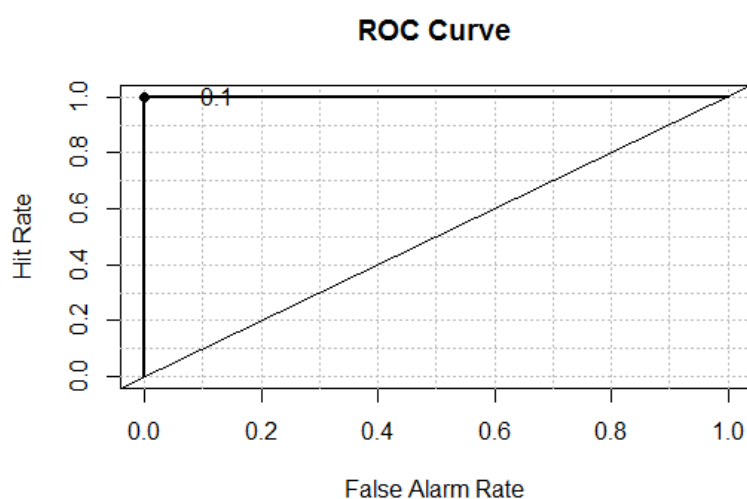
Il modello corrente risulta migliore sia nei confronti del modello saturo (Devianza residua=2.4219, $df=71$) sia nei confronti del modello nullo ($TRV^{oss}=83.9406$, $df=2$, $\alpha^{oss}=0$).

La matrice di confusione (*Tabella 18*) è perfetta: i valori di accuratezza, sensibilità e specificità sono massimi (100%), come fa notare anche la curva ROC in *Figura 17* (AUC=1).

Tabella 18: Matrice di confusione tra la variabile follow-up e i valori predetti dal secondo modello

	Valore predetto=0	Valore predetto=1
Follow-up=0	20	0
Follow-up=1	0	54

Figura 17



Il terzo modello stimato vede come modello iniziale quello che include i tre gruppi (gruppo pervietà, gruppo disfagia e gruppo clinico), per poi arrivare, sempre tramite una *backward elimination*, al modello finale stimato seguente:

$$\log\left(\frac{\hat{\pi}_i}{1 - \hat{\pi}_i}\right) = -4.75 + 8.17 * \text{gruppo pervietà}_i.$$

Tabella 19: Tabella dei coefficienti del terzo modello (nuova procedura)

Parametro	Stima	Standard Error	Z-value	p-value
<i>Intercetta</i>	-4.745	2.482	-1.912	0.056
<i>Gruppo pervietà</i>	8.166	2.598	3.143	0.002

Il modello corrente risulta migliore sia nei confronti del modello saturo (Devianza residua=10.708, *df*=72) sia nei confronti del modello nullo ($TRV^{oss}=75.654$, *df*=1, $\alpha^{oss}=0$).

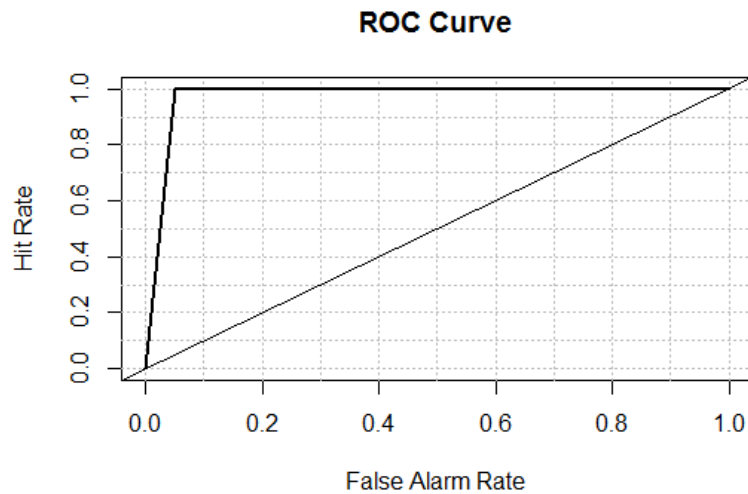
Per valutare l'accuratezza del modello è stata costruita la matrice di confusione (Tabella 20), dalla quale emerge che solo per un soggetto la previsione ottenuta è sbagliata rispetto all'esito della decannulazione. I valori di accuratezza, sensibilità e specificità sono pari, rispettivamente, a 98.6%, 95% e 100%.

Tabella 20: Matrice di confusione tra la variabile follow-up e i valori predetti dal terzo modello

	Valore predetto=0	Valore predetto=1
<i>Follow-up=0</i>	19	1
<i>Follow-up=1</i>	0	54

La stima non parametrica dell'AUC (Figura 18) è pari a 0.975 (I.C.: 0.926-1), sottolineando ancora una volta l'elevata adeguatezza del modello nel prevedere le osservazioni.

Figura 18



Tutti e tre i modelli risultano soddisfacenti nel prevedere in maniera adeguata e accurata l'esito della rimozione della cannula dopo 48 ore. I valori dell'AIC per i tre modelli sono, rispettivamente: 10.77, 6 e 14. Il modello con valore del criterio minimo è il secondo, oltre ad essere quello che prevede correttamente tutti i soggetti. Si tratta anche del modello che si allontana maggiormente da quello nullo.

Secondo questa procedura, nel caso in cui la variabile esito non fosse nota, il confronto tramite AIC e quello tramite il valore del test del log-rapporto di verosimiglianza per il confronto tra il modello corrente e quello nullo portano a preferire il primo modello, ossia quello meno parsimonioso tra i due. Anche il confronto tramite AUC porta a preferire il primo modello.

4.4 Confronto tra i risultati ottenuti

Le due procedure, essendo molto simili, portano ad ottenere sostanzialmente gli stessi risultati. Infatti le covariate che sono state incluse, in quanto significative, in tutte e tre le coppie di modelli (uno ottenuto col metodo di Firth, l'altro con la procedura alternativa) sono le medesime. Si può però notare che le stime dei coefficienti β_j e i relativi *standard errors* risultano leggermente superiori seguendo la seconda procedura di stima, rispetto al calcolo con il metodo di Firth.

Con entrambe le procedure, il modello preferibile è il secondo, ossia quello che include come esplicative le variabili esito e pervietà. La classificazione perfetta a cui porta

questo modello è dovuta alla presenza della variabile esito, la quale assomiglia molto alla variabile risposta considerata. Infatti l'unica differenza tra le due consiste nella durata del test.

Tuttavia, nel caso in cui non fosse possibile conoscere il risultato del test di decannulazione immediatamente dopo la rimozione della cannula (esito), tra i restanti due modelli si arriva, tramite il confronto di AIC, con la prima procedura a preferire quello che include solo la variabile gruppo pervietà, mentre con la seconda procedura risulta preferibile il modello che include le variabili tosse riflessa, chiusura e pervietà. Questa discrepanza è dovuta al fatto che le due procedure applicano una diversa penalizzazione alla funzione di punteggio. La parsimoniosità del terzo modello ($p = 2$) potrebbe far propendere la scelta verso quest'ultimo. Inoltre due variabili esplicative incluse nel primo modello (tosse riflessa e chiusura) sono risultate fortemente dipendenti secondo le analisi svolte nel *Capitolo 2*. Perciò, anche per evitare il rischio di multicollinearità, probabilmente risulta preferibile scegliere tra i due modelli quello che include solo la variabile gruppo pervietà. Si ricorda però che i due modelli oggetto del confronto rispondono ad obiettivi conoscitivi differenti.

Nessuno dei modelli stimati include la variabile tempo di mantenimento della cannula, nonostante la curva ROC presentata nel *Capitolo 3* discriminasse abbastanza bene i pazienti tra coloro che avrebbero superato il test di rimozione della cannula e gli altri. Le variabili incluse in tutti e tre i modelli hanno una relazione significativa con la risposta follow-up, come emerso dalle analisi esplorative presentate nel *Capitolo 2*. In particolare tutti i modelli stimati inseriscono tra le esplicative la variabile pervietà, ad indicare quindi che il diametro del lume gioca un ruolo fondamentale nel determinare la riuscita del tentativo di decannulazione a 48 ore dalla rimozione della cannula. Inoltre la variabile gruppo pervietà, oltre ad includere il risultato del test relativo al diametro del lume, include anche il risultato della chiusura della cannula per 72 ore per testare la capacità del paziente di respirare con naso e bocca. Queste due variabili sono state incluse anche nel primo modello, insieme alla variabile tosse riflessa, ossia la valutazione della reazione del paziente ad un agente che provoca la tosse. Perciò nel momento in cui non sia noto il valore della variabile esito, tutte queste variabili contribuiscono considerevolmente alla determinazione della probabilità di riuscita della decannulazione dopo 48 ore.

CAPITOLO 5

CONCLUSIONI

Nel *Capitolo 1*, dopo una panoramica sull'intervento di tracheostomia e sui relativi rischi, sono state presentate le unità statistiche dello studio, 74 pazienti tracheostomizzati affetti da *Acquired Brain Injury*. Le variabili rilevate sono 14 e riguardano sia caratteristiche demografiche dei pazienti sia il superamento di determinati test. Inoltre sono state costruite tre variabili (i gruppi) tramite combinazione di alcuni test.

L'obiettivo primario di questo studio è rappresentato dal testare la performance dei parametri quantitativi rilevati affinché la decannulazione possa avvenire nel modo più sicuro possibile per i pazienti.

Nel *Capitolo 2* sono state studiate le distribuzioni delle singole variabili, nonché le loro relazioni. È emerso che la maggior parte delle coppie di variabili analizzate presentano una relazione significativa, ma, soprattutto, che la variabile considerata come *golden test*, cioè follow-up, è stocasticamente indipendente solo dalle variabili età, sesso e tosse volontaria.

Sulla base di questi risultati, nel *Capitolo 3* sono stati calcolati alcuni indici sintetici, dai quali è emerso che se si vuole privilegiare la sensibilità, il test diagnostico migliore risulta essere il gruppo clinico, mentre, nel caso in cui si voglia privilegiare la specificità i test migliori sono: chiusura, pervietà, esito e gruppo pervietà. È stata in seguito costruita la curva ROC considerando come variabile criterio il tempo di mantenimento della cannula ed è emerso che i due gruppi di pazienti vengono discriminati abbastanza bene ($AUC=0.812$). Il valore soglia che discrimina questi due gruppi è rappresentato da 178.50 giorni, con una sensibilità, una specificità e un'accuratezza pari, rispettivamente, a 80%, 77.8% e 78.4%.

Infine, nel *Capitolo 4*, sono stati stimati alcuni modelli di regressione logistica allo scopo di poter prevedere nel modo migliore possibile l'esito della decannulazione per i pazienti. Il modello migliore include le variabili esito e pervietà e, con un'area sottesa alla curva ROC pari a 1, prevede correttamente il valore della variabile follow-up per tutti i pazienti. Tuttavia, nel caso in cui la variabile esito non sia nota, il modello

migliore include le variabili tosse riflessa, chiusura e pervietà ($AUC=0.9995$). Per questo modello i valori di sensibilità, specificità e accuratezza sono pari, rispettivamente, a 95%, 100% e 98.6%. Ciò che è stato ottenuto differisce rispetto a quanto emerso dallo studio di Zanata *et al.* (2014), nel quale, invece, erano risultati significativi nel prevedere l'esito della decannulazione: il livello di coscienza, la capacità di tollerare la chiusura della cannula tracheostomica, l'efficacia della tosse e la quantità di secrezione. Infatti solo una di queste variabili (chiusura) è stata inclusa nel modello.

L'obiettivo secondario di questo studio consiste nell'esplorare se da un possibile raggruppamento di alcuni test si possa ottenere una migliore regola predittiva per la rimozione della cannula. Partendo dal modello che include tutte e tre le variabili ricavate dalla combinazione di più test, si è ottenuto un modello che include come esplicativa solamente la variabile gruppo pervietà. Quest'ultima tiene conto sia dell'esito del test relativo alla chiusura della cannula per 72 ore sia della misura del diametro del lume. Questo modello riesce a classificare correttamente il 98.6% dei pazienti, ha una sensibilità e una specificità pari, rispettivamente, a 95% e 100% e presenta un'area sottesa alla curva ROC pari a 0.975.

Tutti i coefficienti delle variabili incluse nei tre diversi modelli hanno segno positivo e ciò implica che l'aver superato almeno uno tra i test usati come esplicative porta ad un incremento del log-rapporto della quota e quindi della probabilità di successo della rimozione della cannula.

Si può notare che le variabili che sono state incluse nei tre modelli sono le stesse che, se utilizzate come test diagnostici (*Capitolo 3*), portavano ad un'elevata specificità (100% per chiusura, pervietà, esito e gruppo pervietà; 98% per tosse riflessa). In questo contesto, infatti, sembra desiderabile una specificità più alta possibile in maniera tale da classificare correttamente tutti i pazienti con follow-up=1.

I tre modelli ottenuti hanno la massima specificità e un'elevata sensibilità. Ciò indica che tutti i pazienti che hanno superato la decannulazione definitiva e la maggior parte dei pazienti che non l'hanno superata vengono previsti in maniera corretta.

In conclusione, poiché le variabili tosse riflessa, chiusura, pervietà e esito giocano un ruolo importante nella determinazione dell'esito della rimozione della cannula, potrebbero rappresentare la base da cui partire nell'identificazione di un protocollo di decannulazione specifico per pazienti con ABI.

BIBLIOGRAFIA

- Armitage, P. e Berry, G. (1996). *Statistica medica: Metodi statistici per la ricerca in Medicina*. McGraw-Hill.
- Azzalini, A. e Scarpa, B. (2004). *Analisi dei dati e data mining*. Springer Verlag.
- Busatto, C. (2016). Tesi di Laurea. *Analisi statistica del test diagnostico sui nervi cranici*. Dipartimento di Scienze Statistiche, Università di Padova.
- D'Arrigo, G., Provenzano, F., Torino, C., Zoccali, C., Tripepi, G. (2011). *I test diagnostici e l'analisi della curva ROC*. Master in epidemiologia clinica. Società italiana di Nefrologia. 28(6): 642-647.
- Duca, P. (1995). *Analisi della curva ROC: significato dell'area ad essa sottesa e sua stima non parametrica*. Statistica Applicata, 7(3): 296-304.
- Durbin, CG Jr. (2010). *Tracheostomy: why, when, and how?* Respir Care, **55**, 1056–1068.
- Firth, D. (1993). *Bias reduction of maximum likelihood estimates*. Biometrika, **80**, 27-38.
- Girardi, A. (2011). Tesi di Laurea. *Metodi per la riduzione della distorsione dello stimatore di massima verosimiglianza*. Dipartimento di Scienze Statistiche, Università di Padova.
- Heinze, G. e Schemper, M. (2002). *A solution to the problem of separation in logistic regression*. Statistics in Medicine, **21**, 2409-2419.
- Kenne Pagui, E.C., Salvan, A., Sartori, N. (2017). *Median bias reduction of maximum likelihood estimates*. Dipartimento di Scienze Statistiche, Università di Padova, ArXiv: 1604.04768v4 [stat.Me].
- Kosmidis, I. (2014). *Bias in parametric estimation: reduction and useful side-effects*. WIRE Computational Statistics, **6**, 185-196.
- Kosmidis, I., Firth, D. (2010). *A generic algorithm for reducing bias in parametric estimation*. Electronic Journal of Statistics, **4**, 1097-1112.
- McCullagh, P. (1987), *Tensor Methods in Statistics*. London: Chapman and Hall.
- Moolgavkar, S. H., Venzon, D. J. (1988). *Method for computing profile-likelihood-based confidence intervals*. Journal of the Royal Statistical Society, Series C: Applied Statistics, 37(1): 87-94.

- Nelder, J.A., Wedderburn, R.W.M. (1972). *Generalized Linear Models*. Journal of the Royal Statistical Society, 135(3): 370-384.
- Pace, L. e Salvan, A. (1996). *Introduzione alla Statistica II - Inferenza, Verosimiglianza, Modelli*. Cedam, Padova.
- Pace, L. e Salvan, A. (1997). *Principles of Statistical Inference from a Neo-Fisherian Perspective*. World Scientific Pub Co Inc., Singapore.
- Quenouille, M.H. (1949). *Approximate tests of correlation in time-series*. J. R. Statist. Soc., **B 11**, 68-84.
- Quenouille, M.H. (1956). *Notes on bias in estimation*. Biometrika, **43**, 353-360.
- Smeltzer, S.C., Bare, B.G., Hinkle, J.L., Kerry, H.C. (2010), *Brunner-Suddarth Infermieristica medico-chirurgica Volume 1*, Casa Editrice Ambrosiana. Unità 5, Capitolo 25.
- Swets, J.A. (1988). *Measuring the accuracy of diagnostic systems*. Science, 240(4857): 1285-1293.
- Zanata, I.L., Santos, R.S., Hirata, G.C. (2014). *Tracheal Decannulation Protocol in Patients Affected by Traumatic Brain Injury*. Thieme Open Access, **18**, 108-114.
- Dispense didattiche di "Statistica Medica". A.A. 2016-2017. A cura della Prof.ssa Laura Ventura.
- Materiale didattico di "Modelli Statistici 2". A.A. 2016-2017. A cura dei Proff. Alessandra Salvan e Nicola Sartori.

SITOGRAFIA

- www.deutera.com/cose-la-disfagia/
- www.uildm.org/scienza-e-medicina/approfondimenti-tematici/tracheostomia-e-comunicazione/
- ospedalesancamillo.net/assistenza/dipartimento-di-neuro-riabilitazione/gravi-cerebrolesioni-acquisite
- http://www.consensusconferencegca.com/assets/files/Documento_preparatorio_CC.pdf
- www.rstudio.com