



POLITECNICO DI TORINO

Corso di Laurea Magistrale in Ingegneria Civile

Tesi di Laurea Magistrale

Analisi dei guasti nella rete acquedottistica di Torino

Relatori:

Prof. Fulvio Boano

Prof. Luca Ridolfi

Ing. Marco Scibetta (SMAT)

Candidato
Mariagrazia Bellantone

Ottobre 2019

Sommario

INTRODUZIONE	1
1 TUBAZIONI PER ACQUEDOTTO, MECCANISMI DI ROTTURA E COSTI AD ESSI LEGATI	4
1.1 TUBAZIONI	4
1.1.1 Tubazioni Metalliche	5
1.1.2 Tubazioni Cementizie	7
1.1.3 Tubazioni Plastiche	8
1.2 MECCANISMI DI ROTTURA	10
1.2.1 Indicatori Ambientali	10
1.2.2 Indicatori Strutturali	11
1.2.3 Indicatori Idraulici	11
1.2.4 Indicatori Operativi	12
1.2.5 Modalità Di Rottura	12
1.3 COSTI LEGATI ALLE ROTTURE	13
1.3.1 Costi Diretti	13
1.3.2 Costi Indiretti	13
2 LA COMPOSIZIONE DELLA RETE IDRICA DI TORINO	14
2.1 CONCLUSIONI	19
3 LA LOGICA DI ACQUISIZIONE DEI DATI	20
3.1 TABELLE FONDAMENTALI	20
3.1.1 Tabella Workorder	21
3.1.2 Tabella Woserviceaddress	22
3.1.3 Tabella Vie	22
3.1.4 Estrazione dei Dati di Interesse e Calcolo delle Occorrenze	23
3.2 TABELLA Wwvcondottetorino	26
3.3 TABELLA ReteTorino	28
3.4 TABELLA RegressioneCondotte	32
3.5 TABELLA RegressioneTGM	33
3.6 CONCLUSIONI	39
4 ANALISI DEI DATI IN FUNZIONE DI FATTORI STRUTTURALI, OPERATIVI E AMBIENTALI	41
4.1 ANALISI DELLA TABELLA ReteTorino	43
4.1.1 Estrazione e analisi di sottotabelle per ogni materiale	50
4.2 ANALISI DELLA TABELLA RegressioneCondotte	69
4.3 ANALISI DELLA TABELLA RegressioneTGM	76
4.3.1 Analisi combinata della variabile 'materiale' e 'anno di posa'	84
4.4 CORRELAZIONI TRA LE VARIABILI INDIPENDENTI	85

4.5	CONCLUSIONI	87
5	LA REGRESSIONE LOGISTICA	88
5.1	IL MODELLO DI REGRESSIONE LOGISTICA	88
5.1.1	La stima dei parametri	88
5.1.2	Funzione Logistica	90
5.1.3	Logit e Odds	90
5.2	COSTRUZIONE DI UN MODELLO DI REGRESSIONE	91
5.3	TEST PER LA BONTÀ DI ADATTAMENTO DEL MODELLO LOGISTICO	92
5.3.1	Il Test Rapporto di Verosimiglianza	93
5.3.2	Lo Pseudo- R^2	94
5.3.3	Test della bontà di adattamento: Devianza, χ^2 di Pearson e Test di Hosmer-Lemeshow	95
5.4	SIGNIFICATIVITÀ DEI PARAMETRI DEL MODELLO	95
5.5	ESEMPIO DIMOSTRATIVO	96
6	APPLICAZIONE DEL MODELLO DI REGRESSIONE LOGISTICA	101
6.1	ESTRAZIONE DI 20 SET DI DATI E STIMA DEL MODELLO	102
6.2	VERIFICA DEI MODELLI APPLICANDO IL METODO BACKWARD	139
6.3	IL MODELLO GENERALE APPLICATO ALLA TABELLA RegressioneTGM	147
6.4	CONCLUSIONI	150
7	ESTENSIONE DEL MODELLO E DEFINIZIONE DEL VALORE DI SOGLIA	152
7.1	ANALISI DELLA VARIABILE 'Tram Route'	152
7.2	APPLICAZIONE DEL MODELLO	154
7.3	DEFINIZIONE DEL VALORE DI SOGLIA	157
7.4	CONCLUSIONI	160
	CONCLUSIONI	161
	APPENDICE A	164
	BIBLIOGRAFIA	193

INTRODUZIONE

Il disciplinamento dell'uso, del controllo e dell'organizzazione delle opere di approvvigionamento acquedottistico e fognario rivestono una funzione determinante in materia di risorse idriche. Trattandosi di un argomento di notevole importanza, questi risultano disciplinati da un fascio di norme alquanto ricco ed articolato, concernenti la tutela dell'ambiente.

Nella gestione dei sistemi acquedottistici, assume un ruolo rilevante il risparmio idrico, diretto sia a limitare qualsivoglia spreco di risorsa da parte del consumatore finale, sia a contenere a valori non anomali le perdite idriche all'interno della rete; quest'ultima, infatti è tutt'oggi caratterizzata da una percentuale media di perdita pari al 39%. La riduzione delle perdite idriche è uno scopo da perseguire, che assicura vantaggi certamente economici, la riduzione dei consumi energetici, ma anche la riduzione dei prelievi di acque superficiali e sotterranee, con conseguenti benefici ambientali.

Le perdite idriche sono riconducibili a più cause e possono essere suddivise in più categorie (consumo non autorizzato, contatori imprecisi, sfiori dai serbatoi, ecc.); la loro riduzione non può prescindere dallo studio delle rotture delle condotte e dall'analisi dei molteplici fattori che determinano la non integrità della rete, che saranno oggetto di studio di questo lavoro di Tesi.

Le rotture delle tubazioni, oltre a causare una considerevole dispersione idrica, richiedono come conseguenza l'interruzione del servizio e creano disagi al consumatore. Il fenomeno in questione ricopre importanza vitale per le aziende, interessate a proporre un'alta qualità del servizio per raggiungere un buon livello di soddisfazione del cliente.

Nel presente lavoro, nato dalla stretta collaborazione tra il *Politecnico di Torino* e la società *SMAT (Società Metropolitana Acque Torino S.p.a.)*, è stata esaminata la problematica relativa alle rotture nel sistema acquedottistico della città di Torino, considerando i guasti riscontratisi nel periodo di osservazione dal 2006 al 2016 ed è stata valutata, in base ai dati a disposizione, la presenza di correlazione tra le rotture in rete, le caratteristiche delle condotte e i fattori operativi ed ambientali, tra questi ultimi vi è il carico dinamico. I valori di traffico sono stati messi a disposizione dalla società *5T*, che opera nel campo dei sistemi ITS (Intelligent Transport Systems) e dell'infomobilità, ed in formato GTFS dall'azienda pubblica di trasporto locale *GTT (Gruppo Torinese Trasporti)*. La tesi segue cronologicamente e logicamente il lavoro svolto in tale ambito dalla collega *Clara Ghigo*, concluso nel dicembre 2017 e dal collega *Giovanni Saporito*, concluso nel settembre 2018. Una corretta previsione delle rotture, permette un'efficace pianificazione degli interventi di manutenzione e rinnovo delle parti danneggiate, prevenendo il guasto e la dispersione di volume idrico, con conseguente beneficio ambientale, energetico e quindi economico. La suddetta analisi si avvarrà di un modello di regressione statistica: il modello di regressione logistica.

La tesi si compone di sette capitoli, si riportano brevemente informazioni sulla trattazione di ciascun capitolo:

1. **Capitolo 1.** Il capitolo descrive i materiali che vengono adoperati per la costruzione delle tubazioni nei sistemi acquedottistici e vengono esposti i fattori che causano la rottura delle tubazioni stesse.
2. **Capitolo 2.** Nel capitolo viene analizzata la composizione della rete idrica del comune di Torino gestita dal gruppo *SMAT* (*Società Metropolitana Acque Torino*).
3. **Capitolo 3.** Il capitolo dettaglia la modalità di raccolta dati da parte di *SMAT* per la coordinazione degli interventi operati nei 289 comuni gestiti dalla società. Saranno esposte le tabelle, contenenti le informazioni sui guasti, ed in particolare gli interventi “*fuga condotta*”. L’unione di queste tabelle darà vita alla tabella *Wwvcondottetorino*; l’intersezione con il database cartografico consentirà di ottenere dapprima la tabella *ReteTorino*, contenente gli elementi condotta con le proprie caratteristiche, di cui alcuni privi di informazioni in tutti i campi ed in seguito ad operazioni di filtraggio e integrazione di caratteristiche, si otterrà la tabella *RegressioneCondotte*. Infine l’intersezione di quest’ultima con i dati sul servizio *GTT* (*Gruppo Torinese Trasporti*), il TGM (*Traffico Giornaliero Medio*) e le velocità medie stimate sulle strade presenti nel modello del Supervisore Metropolitano della società *5T*, ha consentito di ottenere la tabella *RegressioneTGM*. Quest’ultima sarà il punto di partenza per la stima del modello di regressione e la susseguente fase di testing dello stesso. Purtroppo nella costruzione di tale tabella, verranno escluse parecchie condotte, ma questa operazione sarà necessaria per poter basare la regressione esclusivamente sui tratti aventi tutte le informazioni utili.
4. **Capitolo 4.** Contiene l’analisi dettagliata delle tabelle ricavate e la conseguente valutazione delle classi di condotte più vulnerabili. La stima è stata eseguita analizzando una serie di fattori ritenuti possibili responsabili delle rotture, classificabili in tre categorie: fattori fisici-strutturali, fattori operativi e fattori ambientali.
5. **Capitolo 5.** Nel capitolo viene illustrato il modello di regressione logistica e come questo sia in grado di determinare la forza di associazione fra un fattore di rischio e la variabile dipendente, ma anche la probabilità di realizzarsi dell’evento.
6. **Capitolo 6.** Vengono riportati i risultati dell’applicazione del modello di regressione logistica alla tabella *RegressioneTGM*. Vengono ricavati 20 set di dati, estratti dalla tabella precedente in modo da mantenere inalterata la percentuale di rottura della tabella *ReteTorino*. Per comprendere le prestazioni del modello, ogni set di dati è diviso in un set di allenamento, sul quale viene creato il modello, e un set di test, sul quale avviene il fit e la previsione del modello. Il modello stima valori di probabilità media di rottura maggiori per condotte realmente rotte, sia applicando i coefficienti di regressione stimati sui 20 set di fitting e applicati sui 20 set di testing, sia applicando il modello sulla tabella *RegressioneTGM*.

7. **Capitolo 7.** Il capitolo presenta i risultati dell'applicazione del modello di regressione logistica alla tabella *RegressioneCondotte*, a cui viene aggiunta l'informazione sulla presenza del tram. Verrà quindi definita un valore di cut-off, oltre la quale considerare una condotta suscettibile di rottura.

1 TUBAZIONI PER ACQUEDOTTO, MECCANISMI DI ROTTURA E COSTI AD ESSI LEGATI

Il termine acquedotto ha etimologia latina, derivando da “*aquaeductus*”, cioè condurre acqua. Quindi, secondo la detta accezione etimologica, è da definire acquedotto qualunque opera adatta al trasporto dell’acqua; perciò rientrerebbero negli acquedotti anche le condotte a scopo irriguo, di produzione di energia elettrica e le stesse fognature. Invero l’accezione più comune, che è poi quella consolidatasi nel tempo è ben diversa. Infatti con acquedotto si intende il complesso delle opere per la captazione, il trasporto, lo stoccaggio e la distribuzione delle acque intese come materie prime (*Mantica, 1979*).

1.1 TUBAZIONI

L’alimentazione delle utenze in una rete di distribuzione idrica viene realizzata tramite condotte in pressione, che costituiscono quindi uno degli elementi di maggiore importanza di un acquedotto.

I parametri che consentono di identificare le condotte sono:

- ❖ la pressione nominale PN, il cui valore finale è costituito dalla somma di¹:
 - “La pressione di esercizio P_e , intesa come massima pressione assiale conseguente ai possibili prevedibili esercizi del sistema idraulico, ivi comprese le massime sovrappressioni di moto vario conseguenti a prevedibili condizioni di esercizio, quando anche di carattere temporaneo e/o accidentale”;
 - “La pressione equivalente P_o , pressione che trasmette al tubo tensioni di trazione massima uguali a quelle calcolate per condizioni non connesse con l’esercizio idraulico del sistema (come carico del terreno, sovraccarico accidentale, condizioni di appoggio, variazioni termiche, azioni sismiche, ecc.) e, nei tubi di grande diametro, a quelle conseguenti alla distribuzione idrostatica delle pressioni sul diametro verticale”.
- ❖ il diametro nominale DN, con cui si definisce convenzionalmente la dimensione delle condotte. Ogni produttore determina tale grandezza seguendo una propria convenzione, infatti il diametro nominale non coincide in genere né con il diametro interno né con il diametro esterno. Poiché le dimensioni delle tubazioni condizionano velocità e perdite di

¹ Decreto Ministero LL.PP. 12 Dicembre 1985: “*Norme tecniche relative alle tubazioni*”.

carico, è bene verificare meticolosamente i calcoli idraulici in funzione del valore del diametro interno.

Vengono descritte le tipologie di materiali ordinariamente adoperate in ambito acquedottistico; includendo sia materiali di recente diffusione ed utilizzo (materiali plastici), sia materiali non più commerciabili sul mercato (cemento amianto).

❖ Tubazioni metalliche:

- ghisa grigia;
- ghisa sferoidale;
- acciaio;
- chameroy.

❖ Tubazioni cementizie:

- cemento armato;
- cemento armato precompresso;
- cemento amianto.

❖ Tubazioni plastiche:

- policloruro di vinile (PVC);
- polietilene a bassa densità (PEBD);
- polietilene ad alta densità (PEAD);
- poliestere rinforzato con fibre di vetro (PRFV).

1.1.1 Tubazioni Metalliche

1.1.1.1 Ghisa Grigia

Il materiale venne ampiamente utilizzato fino agli anni '50 per la costruzione di acquedotti, in quanto era difficilmente corrodibile dall'acqua e dall'ambiente esterno. La ghisa grigia è costituita da una percentuale elevata di carbonio e deve il suo nome alla disposizione dei cristalli di grafite.

La sua composizione la rende molto lavorabile, ma altamente fragile; le tubazioni sono ottenute per colata verticale della ghisa d'alto forno e ghisa ordinaria fuse.

Lo sviluppo e l'introduzione di materiali con caratteristiche meccaniche più elevate hanno limitato nel tempo l'impiego della ghisa grigia, che rimane però tra i materiali attualmente più presenti nelle reti.

1.1.1.2 Ghisa Sferoidale

È ottenuta aggiungendo una limitata percentuale di magnesio alla ghisa grigia fusa. A differenza della precedente, i cristalli di grafite nella ghisa sferoidale si presentano sotto forma di noduli o micro-sfere. Il magnesio e la disposizione della grafite le conferiscono una struttura molecolare più stabile e conseguentemente migliori caratteristiche meccaniche in

termini di resistenza a trazione e riduzione del comportamento fragile. A livello operativo il materiale ben si presta all'assorbimento delle variazioni e resiste all'uso prolungato senza deteriorarsi e corrodersi.

I requisiti ed i relativi metodi di prova applicabili ai tubi, raccordi ed accessori di ghisa sferoidale destinati alla costruzione di condotte fuori dagli edifici e quindi la produzione stessa sono normati a livello europeo².

1.1.1.3 Acciaio

Le tubazioni in acciaio sono largamente utilizzate per l'adduzione d'acqua potabile in virtù della caratteristica di igienicità. Il materiale, inoltre, possiede elevata resistenza meccanica e il vantaggio di poter realizzare separatamente più pezzi, saldandoli successivamente.

Anche per l'acciaio è prevista una normativa europea che definisce i controlli e le prove da effettuare sia sulle condotte, che sui raccordi; in particolare le norme prevedono una serie di prove sul materiale, sul prodotto finito e sulle saldature³.

Le condotte in acciaio sono però soggette alla corrosione se posate su terreni aggressivi. Un ruolo fondamentale per la protezione dalla corrosione, assumono i rivestimenti sia esterni che interni. Il rivestimento esterno isola le condotte dall'ambiente e dagli agenti esterni, quelli più comunemente utilizzati sono quelli bituminosi⁴ e quelli in polietilene fuso o estruso⁵. I rivestimenti interni proteggono le condotte dai fenomeni corrosivi e migliorano le prestazioni idrauliche; nel caso di tubazioni per acquedotto vengono utilizzate vernici epossidiche, rivestimento cementizio e zincatura. In alternativa per acque chimicamente aggressive, si utilizza la bitumatura.

1.1.1.4 Chameroy

I tubi Chameroy non sono che tubi di lamiera di ferro stagnata e chiodata, rivestiti all'esterno e difesi dall'ossidazione con catrame, fasciati di tela ed ancora spalmati di bitume; internamente hanno una vernice dura ed inalterabile all'acqua, ma fragile. Devono il loro nome a M. Chameroy, che a Parigi nel 1867 li brevettò e iniziò la produzione nella fabbrica omonima. Con il tempo tale materiale cadde in disuso a causa della mancanza di rigidità, che non preveniva il collasso.

² Norma UNI EN 545:2010: *"Tubi, raccordi e accessori di ghisa sferoidale e loro assemblaggi per condotte d'acqua – Requisiti e metodi di prova"*.

³ Norma (UNI, 2019) EN 10020:2001: *"Definizione e classificazione dei tipi di acciaio"*; Norma UNI EN 10224:2006: *"Tubi e raccordi di acciaio non legato per il convogliamento di acqua e di altri liquidi acquosi – Condizioni tecniche di fornitura"*.

⁴ Norma UNI ISO 5256:1987: *"tubi ed accessori di acciaio impiegati per tubazioni interrate o immerse. Rivestimento esterno e interno a base di bitume o di catrame"*.

⁵ Norma UNI 9099:1989: *"Tubi di acciaio impiegati per tubazioni interrate o sommerse. Rivestimento esterno di polietilene applicato per estrusione"*.

1.1.2 Tubazioni Cementizie

Il calcestruzzo presenta bassa resistenza a trazione ed è un materiale naturalmente fragile; queste qualità lo rendono inadatto per la costruzione di condotte soggette anche a modeste pressioni.

Nel tempo sono stati adottati sistemi di miglioramento, che possono essere riassunti nell'introduzione nel materiale di armature metalliche, nell'utilizzo di tecniche di vibrazione durante la messa in opera e nell'introduzione di uno stato tensionale artificialmente indotto, atto a sopperire alla scarsa resistenza a trazione del conglomerato cementizio, ossia la precompressione.

Le migliorie elencate, affinandone le qualità, fanno sì che, a partire dal secondo dopoguerra, venga incrementata la produzione delle tubazioni cementizie.

1.1.2.1 *Cemento Armato e Cemento Armato Precompresso*

All'interno del calcestruzzo è immersa l'armatura, che è costituita da una gabbia elicoidale in ferro. Questa interagisce con il calcestruzzo e assicura maggiore resistenza flessionale e maggiore capacità di assorbire gli sforzi di trazione prodotti dalla pressione interna dell'acqua. A livello chimico si osserva che l'armatura viene passivata e protetta dalle infiltrazioni d'acqua dal conglomerato cementizio, che crea un ambiente basico.

L'aumento di spessore, dovuto all'inserimento dell'armatura, rende necessaria l'applicazione della precompressione, che assolve il compito di ridurre gli stati di tensione o l'applicazione di un rivestimento esterno, composto da mastice bituminoso rinforzato con tessuto di vetro, per salvaguardare l'armatura più vicina alla superficie esterna dalla corrosione. Le condotte in cemento garantiscono lunga durata e stabilità delle caratteristiche idrauliche e hanno bassi costi di produzione, ma il peso notevole aumenta i costi di trasporto.

La condotta per essere immessa sul mercato deve superare una serie di prove da effettuarsi in stabilimento⁶ ed infine l'opera dovrà essere sottoposta alle prove di collaudo facendo riferimento alle norme tecniche per l'esecuzione delle opere in cemento armato precompresso.

1.1.2.2 *Cemento Amianto*

I tubi in cemento amianto sono costituiti da un minerale naturale di consistenza fibrosa misto a cemento. L'amianto è stato in passato ampiamente utilizzato per la buona resistenza al fuoco, per le proprietà isolanti, per la facilità di lavorazione ed anche per il basso costo.

⁶ Decreto Ministero LL.PP. 12 Dicembre 1985: "Norme tecniche relative alle tubazioni".

La sua scoperta è legata allo sviluppo dello sfruttamento minerario degli anni della prima Rivoluzione Industriale. La licenza fu affidata all'Ing. Adolfo Mazza della Eternit S.p.a.. Nel 1992, l'Italia è diventata il primo paese europeo ad introdurre il bando completo dell'amianto⁷, riconoscendone la pericolosità e sancendo il divieto dell'estrazione, importazione, esportazione, commercializzazione e produzione dei prodotti. La potenziale nocività dell'amianto dipende dall'eventualità che siano rilasciate fibre aerodisperse nell'ambiente, che possono essere inalate dagli individui. Le tubazioni in cemento amianto non rappresentano un pericolo per la salute pubblica se integre ed interrate, ma creano condizioni di rischio se oggetto di interventi di manutenzione o sostituzione. La comunità scientifica internazionale ritiene che non vi siano rischi legati all'ingestione di fibre contenute nell'acqua che scorre all'interno delle condotte.

1.1.3 Tubazioni Plastiche

La progressiva diffusione delle tubazioni in materiale plastico è avvenuta in tempi recenti, per cui non si hanno ancora dati reali che riescano a descrivere il loro comportamento a lungo termine. Le proprietà vantaggiose offerte da queste tubazioni sono il peso contenuto, che riduce i costi di trasporto e posa in opera; l'elevata resistenza alla corrosione, all'aggressività dell'acqua e l'elevato isolamento elettrico; la scabrezza idraulica ridotta e la sua costanza nel tempo; l'assoluta impermeabilità. Di contro le tubazioni plastiche presentano limitata resistenza meccanica, sono molto deformabili sotto i carichi esterni, e presentano un decadimento delle caratteristiche meccaniche durante l'invecchiamento e con variazioni di temperatura.

Le tubazioni in plastica sono sostanzialmente costituiti da due materiali: le resine termoplastiche (policloruro di vinile PVC, polietilene a bassa densità PEBD, polietilene ad alta densità PEAD) e le resine termoindurenti rinforzate (resina termoindurente rinforzato con fibre di vetro PRFV). Le differenze fra i due materiali, oltre la risposta correlata alla resistenza meccanica, possono essere individuate nel grado di deformabilità e nella reazione alla temperatura. Mentre le resine termoplastiche rammolliscono se vengono scaldate, le termoindurenti mantengono intatte le caratteristiche acquisite.

1.1.3.1 Policloruro di Vinile – PVC

Il PVC è una resina termoplastica prodotta dalla polimerizzazione del cloruro di vinile, che rammollisce con il calore e conserva, una volta raffreddato la forma impressagli nella fase di rammollimento. I tubi vengono prodotti per estrusione a caldo e sono pressoché atossici se costituiti da materiale puro, ma i catalizzatori e gli additivi utilizzati per contrastare le proprietà negative, possono non esserlo; per questo motivo le tubazioni di materiale plastico

⁷ Legge n. 257 del 27/03/1992: “*Norme relative alla cessazione dell'impiego dell'amianto*” (aggiornata con le modifiche apportate dalla legge 24 aprile 1998, n. 128, della legge 9 dicembre 1998, n. 426, dal decreto-legge 5 giugno 1993, n. 169 e dal decreto-legge 1 ottobre 1996, n. 510).

sono più frequentemente utilizzate nella realizzazione di fognature, mentre se utilizzate per l'adduzione d'acqua potabile devono rispettare oltre le articolate norme tecniche⁸, anche le prescrizioni del Ministero della Sanità.

1.1.3.2 Polietilene a Bassa Densità – PEBD

Il PEBD è un polimero termoplastico ricavato dal petrolio. È inodore ed insapore, e l'atossicità lo rende appropriato per utilizzi che implicano il contatto con alimenti. È un materiale viscoso, che presenta ridotta resistenza a trazione e resistenza chimica, motivo per cui deve essere sottoposto a varie prove di tenuta e resistenza⁹. Il PEBD è più ampiamente utilizzato per la fabbricazione di flaconi ed elementi di imballaggio, e limitatamente per condotte in pressione.

1.1.3.3 Polietilene ad Alta Densità – PEAD

Il polietilene ad alta densità è prodotto dalla polimerizzazione dell'etilene a pressioni e temperature inferiori da quelle richieste dal processo di produzione del PEBD. Anche il PEAD è un polimero insapore, inodore e atossico e quindi ideale per il contatto con l'acqua potabile e gli alimenti. Rispetto al grado a bassa densità presenta una trazione più elevata, una maggiore temperatura di fusione ed una maggiore resistenza chimica. L'alta resistenza a molti solventi, lo rende adatto ad una grande varietà di applicazioni.

La norma che specifica gli aspetti generali dei sistemi di tubazioni di PE è la UNI EN 12201¹⁰.

1.1.3.4 Poliestere Rinforzato con Fibre di Vetro – PRFV

Le tubazioni di PRFV sono costituite da una matrice di resine termoindurenti rinforzate con fibre di vetro. Si individuano non meno di tre strati dalle differenti funzioni. L'interno, a contatto con il fluido, garantisce elevate prestazioni idrauliche, elevata resistenza chimica ed impermeabilità. Lo strato intermedio svolge una funzione meccanico resistente. Lo strato esterno di finitura, protegge le fibre sottostanti dall'attacco ambientale. Il PRFV mostra comportamento visco-elastico, ma meno accentuato in confronto ai materiali termoplastici esaminati. Le caratteristiche meccaniche decadono significativamente solo in presenza di temperature molto elevate ($t > 80^{\circ}\text{C}$), pertanto la durata tecnica del materiale risulta molto lunga. Hanno buona resistenza all'urto e all'abrasione.

⁸ Norma UNI EN ISO 1452:2010: "Sistemi di tubazioni di materia plastica per adduzione d'acqua e per fognature e scarichi interrati e fuori terra in pressione - Policloruro di vinile non plastificato (PVC-U)" e parti successive.

⁹ Norma UNI 7990:2015: "Tubi di polietilene a bassa densità – Dimensioni, requisiti e metodi di prova".

¹⁰ Norma UNI EN 12201:2012: "Sistemi di tubazioni di materia plastica per la distribuzione dell'acqua, e per scarico e fognature in pressione - Polietilene (PE)" e parti successive.

Il materiale è stato immesso sul mercato intorno gli anni '50, ma la normativa che li tratta è piuttosto recente e ha lo scopo di indirizzare nella selezione dei requisiti normativi delle tubazioni di PRFV da installare in relazione alle caratteristiche del fluido da convogliare e alle condizioni di installazione¹¹.

1.2 MECCANISMI DI ROTTURA

Le opere di collegamento tra gli impianti e le utenze dei fruitori, rappresentano una notevole sfida di gestione. La qualità della rete è funzione del suo deterioramento nel tempo, sia strutturale che fisico, e all'aumentare di questo, si registra un aumento del tasso di rottura, una riduzione della capacità idraulica della rete, nonché un abbassamento della qualità dell'acqua nel sistema di distribuzione. L'evoluzione temporale del guasto è influenzata da fattori intrinseci proprie delle tubazioni e non meno determinante è l'influenza dell'ambiente circostante.

Il deterioramento dei tubi può essere classificato in due categorie (*Kleiner & Rajani, 2002*). Il primo è un deterioramento definibile di natura strutturale, che causa la diminuzione della resistenza a rottura delle tubazioni e quindi la capacità di sopportare lo stress indotto da carichi esterni. Il secondo è un deterioramento della superficie interna del tubo, che causa un calo della capacità idraulica e se esternamente è compresente il fenomeno della corrosione in stato avanzato, si riscontra la riduzione di resistenza a rottura.

La rottura del tubo con molta probabilità si verifica quando l'integrità strutturale del tubo viene compromessa dalle sollecitazioni ambientali e operative. Globalmente, le variabili più importanti che descrivono il deterioramento strutturale delle reti idriche possono essere raggruppate in quattro categorie (*Aslani, 2003*):

- ❖ indicatori ambientali;
- ❖ indicatori strutturali;
- ❖ indicatori idraulici;
- ❖ indicatori operativi.

1.2.1 Indicatori Ambientali

La rottura del tubo è frequentemente causata dal situ in cui la condotta è ubicata, esponendola a carichi indotti dal traffico, a variazioni climatiche e a movimenti del terreno. Questi ultimi possono provocare sollecitazioni sulla condotta che conducono alla rottura, così come la naturale corrosività e quindi la natura aggressiva del terreno può provocare la frattura della tubazione in assenza di un rivestimento adeguato. La ricerca ha confermato

¹¹ Norma UNI 9032:2008: "Tubazioni di resine termoindurenti rinforzate con fibre di vetro (PRFV) con o senza cariche – Linee guida per la definizione dei requisiti per l'impiego".

che le condotte interrate sono suscettibili di rottura anche se soggette a condizioni di stress causate dalle vibrazioni indotte dal transito dei mezzi.

Con l'approssimarsi della stagione invernale e più in generale in concomitanza di notevoli variazioni di temperatura, si registra un'elevata frequenza di rottura, che è stata attribuita ai carichi da gelo, ossia un accrescimento dei carichi del terreno, che vengono parzialmente trasferiti alle tubazioni. La riduzione delle temperature fa sì che il gelo penetri più in profondità nel terreno, superando la profondità di posa della condotta e determinando un abbassamento di temperatura della condotta stessa; questo provoca lo sviluppo di tensioni di trazione sul tubo e allo stesso tempo il terreno di confinamento impedisce le deformazioni. Nelle condotte che presentano difetti può facilmente verificarsi che il processo descritto inneschi il fenomeno della rottura, in quanto le condotte risultano più esposte a collasso per trazione.

1.2.2 Indicatori Strutturali

“La rottura di una condotta è direttamente correlata alla sua condizione fisica/strutturale” (*Stephens et al.*, 2003). La letteratura considera tra gli indicatori strutturali l'anno di posa della condotta e quindi gli anni di esercizio. Tale parametro non giustifica, se preso singolarmente, la necessità di intervento e manutenzione, ma è doveroso valutare anche altri fattori. Tra gli indicatori è annoverato il numero di rotture precedenti, che influenza l'insorgere di una nuova rottura, difatti i modelli di regressione predicono probabilità di rottura crescenti per condotte che abbiano presentato più di un guasto.

1.2.3 Indicatori Idraulici

Una tubazione interrata è progettata per resistere a forze interne, come la pressione esercitata dall'acqua, e a forze esterne, come la spinta del terreno. La resistenza della tubazione è offerta dallo spessore della parete e dalla capacità del materiale, che la costituisce, di sopportare sforzi di trazione. Poiché le condotte si deteriorano nel tempo, la loro resistenza decade, rendendole suscettibili di rottura se i carichi applicati fanno sì che venga superato il valore limite di resistenza.

In condizioni di esercizio, i fenomeni oscillatori di pressione, determinati dalle variazioni di velocità del flusso dell'acqua, implicano un regime non costante; la stabilità delle tubazioni, che operano in condizioni instabili, può essere compromessa.

La tipologia di materiale, combinata con la natura più o meno aggressiva dell'acqua, espongono la condotta al fenomeno della corrosione interna, che riduce la sezione trasversale, non garantendo il deflusso di progetto, e riduce lo spessore della parete, aumentando il rischio di rottura.

1.2.4 Indicatori Operativi

Gli aspetti operativi includono errori umani, prodotti non conformi agli standard di progetto, insufficiente controlli di qualità in fabbricazione. La presenza di anomalie preesistenti o insorte in seguito a stoccaggi e depositi non opportuni, come l'esposizione prolungata ai raggi ultravioletti (UV) o altro tipo di agenti atmosferici, riducono la resistenza della condotta, generando rottura. Inoltre, se le condotte non sono caricate e disposte adeguatamente sui mezzi di trasporto, possono verificarsi crepe a causa di un fenomeno chiamato *transit fatigue*. Le prove di collaudo, eseguite prima che la condotta e più in generale il sistema, in cui essa è inserita, possano essere messe in servizio, consentono di individuare tali difetti.

Dalla fase di progettazione, a quella di costruzione, messa in posa e, successivamente, durante la fase di manutenzione, possono verificarsi condizioni che contribuiscono alla rottura della tubazione.

Oltre i fattori descritti, responsabili del meccanismo di rottura nelle tubazioni, ve ne sono degli altri che andrebbero ricercati nelle modalità di gestione aziendale in termini di efficacia ed efficienza e nel piano di manutenzione.

1.2.5 Modalità Di Rottura

Le tipologie di rottura possono essere catalogate come:

- ❖ rotture circonferenziali, rappresentano la condizione più frequente di fallimento per le condotte di diametro piccolo e sono provocate da sollecitazioni longitudinali. Sono prodotte da fenomeni quali: la contrazione e dilatazione termica, le sollecitazioni di flessione, la realizzazione della trincea e del letto di posa con tecniche non adeguate;
- ❖ rotture longitudinali, provocate da sollecitazioni trasversali (circonferenziali) e sono più comuni nei tubi di grande diametro. Le rotture longitudinali interessano per buona parte o per intero la lunghezza dei tubi. Sono dovute alla pressione nel tubo e a quella indotta dal terreno, utilizzato per il riempito dello scavo, e dal traffico, artefice dei carichi dinamici;
- ❖ split bell, provocato da sollecitazioni, agenti sul giunto, di tipo trasversale. La rottura del giunto è diffusa nelle reti acquedottistiche, questa avviene attraverso l'azione combinata degli agenti che innescano la corrosione, del movimento del terreno, delle pratiche di costruzione non eseguite a regola d'arte, della pressione dell'acqua.

In qualsiasi caso, al fine di riabilitare, mediante riparazione o ripristino, la condotta è importante che sia individuata in maniera opportuna la matrice dell'imperfezione.

1.3 COSTI LEGATI ALLE ROTTURE

Le reti esistenti necessitano di continue manutenzioni per assicurare funzionalità, efficienza ed economicità; pertanto vanno costantemente ispezionate e risanate. Gli interventi comportano notevoli costi economici, che si configurano in congestione, ritardi, disservizi, danni alla strada, rifacimento delle opere presenti; costi ambientali, quali rumori ed emissioni; costi sociali tra cui l'interruzione del traffico e delle altre attività produttive connesse, disagi, insalubrità. Negli interventi di manutenzione delle condotte è necessario quindi operare cercando di ridurre non solo i costi di costruzione, definiti costi diretti, ma anche i costi sociali ed ambientali, definiti costi indiretti.

1.3.1 Costi Diretti

I costi diretti sono quelli sostenuti per la costruzione dell'opera e sono quantificabili, in quanto dipendono dal tracciato e dalla tecnica di costruzione impiegata. Questi riguardano oltre la riparazione della rottura, la quantificazione dell'acqua persa e la valutazione dei danni diretti alle proprietà.

1.3.2 Costi Indiretti

I costi indiretti sono quelli non legati direttamente all'esecuzione delle opere, ma che condizionano le attività quotidiane dei cittadini. Sono i costi che derivano dall'impatto negativo dell'intervento sull'ambiente e sulle attività umane e in definitiva sostenuti dalla collettività. Questi costi sono di difficile determinazione e per tale ragione quasi mai contabilizzati.

2 LA COMPOSIZIONE DELLA RETE IDRICA DI TORINO

Il gruppo *SMAT* garantisce l'erogazione del Servizio Idrico Integrato in 289 comuni del torinese, corrispondenti al 95,4% dei 303 comuni totali, per 2.247.449 abitanti. Oggetto di studio del presente lavoro è la rete acquedottistica del comune di Torino. L'estensione della rete e i limiti comunali sono riportati in Figura 2-1.



Figura 2-1 Rappresentazione dell'estensione della rete di distribuzione di Torino

La rete di distribuzione di Torino si espande per 1640 chilometri ed è costituita da 154.725 tubazioni. Mediante la rete di distribuzione, *SMAT* eroga una portata media giornaliera di 7.000 litri, nel giorno di massimo consumo, per il fabbisogno. Viene sviluppata una descrizione dello stato strutturale della rete attraverso l'elaborazione dei dati ottenuti dal database cartografico della società *SMAT*. È stata posta l'attenzione sulle seguenti caratteristiche delle condotte:

- ❖ ubicazione geografica;
- ❖ lunghezza;

- ❖ materiale;
- ❖ diametro;
- ❖ anno di posa.

Ogni tipologia di dato disponibile è stato analizzato in termini di numero di condotte ed estensione chilometrica, individuando delle classi di valori. Sono state calcolate le percentuali di appartenenza alle classi rispetto al totale per poter disporre di una più immediata visione delle caratteristiche delle condotte.

Lunghezza delle condotte

Individuate cinque classi di lunghezza delle condotte, i risultati ottenuti sono riportati in Tabella 2-1.

Tabella 2-1 Analisi condotte suddivise in classi di lunghezza

-	Lunghezza [m]	N° Condotte	%	Chilometri	%
CLASSE 1:	0 m – 10 m	129046	83.4	134.2	8.2
CLASSE 2:	10 m – 50 m	14558	9.4	385.62	23.6
CLASSE 3:	50 m – 100 m	7536	4.9	533.57	32.7
CLASSE 4:	100 m – 300 m	3365	2.2	496	30.4
CLASSE 5:	300 m – 600 m	220	0.1	84.2	5.2
	TOT	154725	100.0	1633.59	100.0

È evidente che le condotte aventi lunghezza inferiore a 10 metri rappresentano la quasi totalità della rete in termini di numero condotte, ma solo l'8.2% della lunghezza totale della rete. Infatti circa il 60% della rete è caratterizzato da condotte comprese tra i 50 e i 300 metri.

La copertura chilometrica delle cinque classi è riportata nel grafico in Figura 2-2.

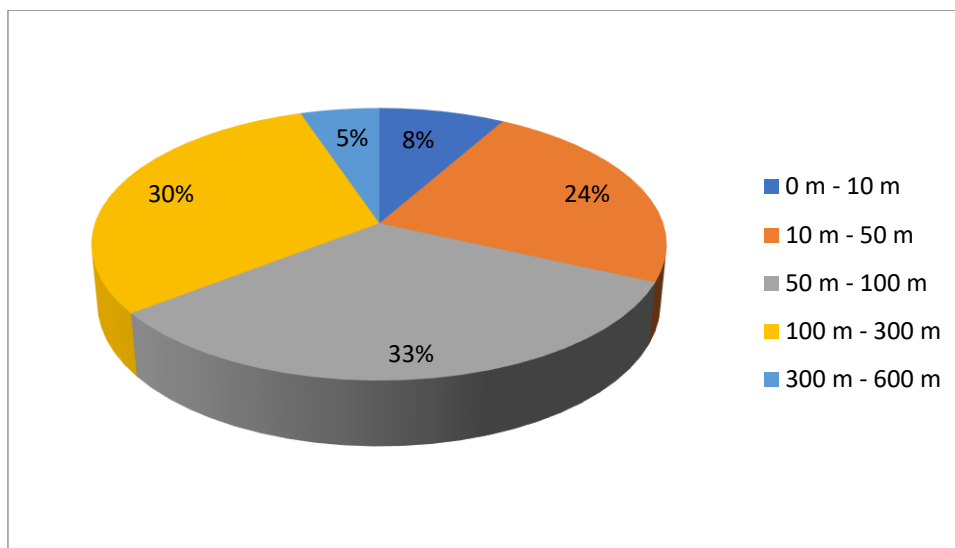


Figura 2-2 Suddivisione delle condotte in classi di lunghezza e rappresentazione delle percentuali

Materiale

Il materiale è una caratteristica che, come già esposto nel Cap. 1, influenza notevolmente le prestazioni delle condotte e la possibile insorgenza di rotture. La suddivisione delle condotte per classi di materiale, le percentuali di appartenenza in funzione del numero di condotte e i chilometri presenti in rete sono riportati in Tabella 2-2.

Tabella 2-2 Analisi condotte suddivise in classi di materiale

Materiale	N° Condotte	%	Chilometri	%
Ghisa grigia	79545	51.4	950.8	58.1
Ghisa sferoidale	31334	20.3	417.2	25.5
Acciaio	15409	10.0	180.9	11.1
Eternit	2775	1.8	53.9	3.3
Materiali Indefiniti	24394	15.8	11.9	0.7
PEAD	754	0.5	8.9	0.5
Chameroy	32	0.0	4.9	0.3
PVC	248	0.2	4.1	0.2
Ferro	200	0.1	1.8	0.1
Cemento armato	7	0.0	0.7	0.0
Piombo	24	0.0	0.5	0.0
Pebd	6	0.0	0.1	0.0
TOT	154728	100.0	1635.6	100.0

Dalla Tabella 2-2 è possibile notare che l'utilizzo della ghisa grigia è ampiamente diffuso all'interno della rete, sia in termini di numero condotte che di lunghezza totale, seguito dalla ghisa sferoidale, che ha nel tempo sostituito la prima. Per alcuni materiali, nonostante

presentino un numero abbastanza elevato di condotte, la percentuale chilometrica in rete risulta essere bassa. La Figura 2-3 dettaglia l'andamento delle percentuali.

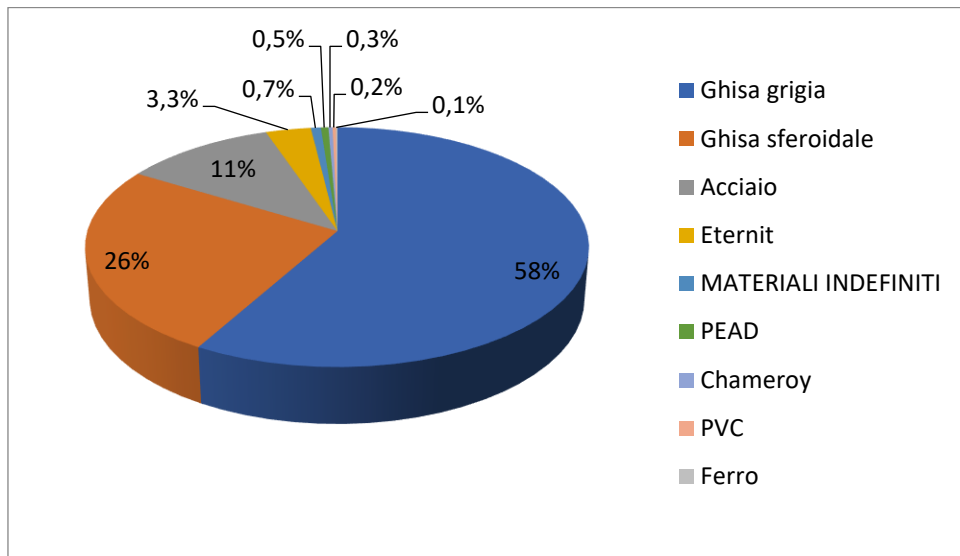


Figura 2-3 Suddivisione delle condotte in classi di materiali e rappresentazione delle percentuali

Diametro

Per una visione immediata dei diametri che costituiscono la rete, sono state considerate tre classi di riferimento; i dati ottenuti sono riportati in Tabella 2-3.

Tabella 2-3 Analisi condotte suddivise in classi di diametri

Diametro		N° condotte	%	Chilometri	%
CLASSE 1:	D ≤ 100 mm	105375	68.1	888.5	54.3
CLASSE 2:	100mm ≤ D ≤ 200 mm	36383	23.5	411.2	25.1
CLASSE 3:	D ≥ 200 mm	12964	8.4	335.9	20.5
	TOT	154722	100.0	1635.6	100.0

La tabella mostra che oltre la metà dei diametri presenta dimensione inferiore a 100 millimetri. Un'altra informazione riguarda la lunghezza media delle condotte, infatti dai risultati si evince che a maggiori diametri, corrispondono maggiori lunghezze; le condotte di diametro maggiore di 200 millimetri risultano essere meno di 13000, ma costituiscono più del 20% in lunghezza della rete (Figura 2-4).

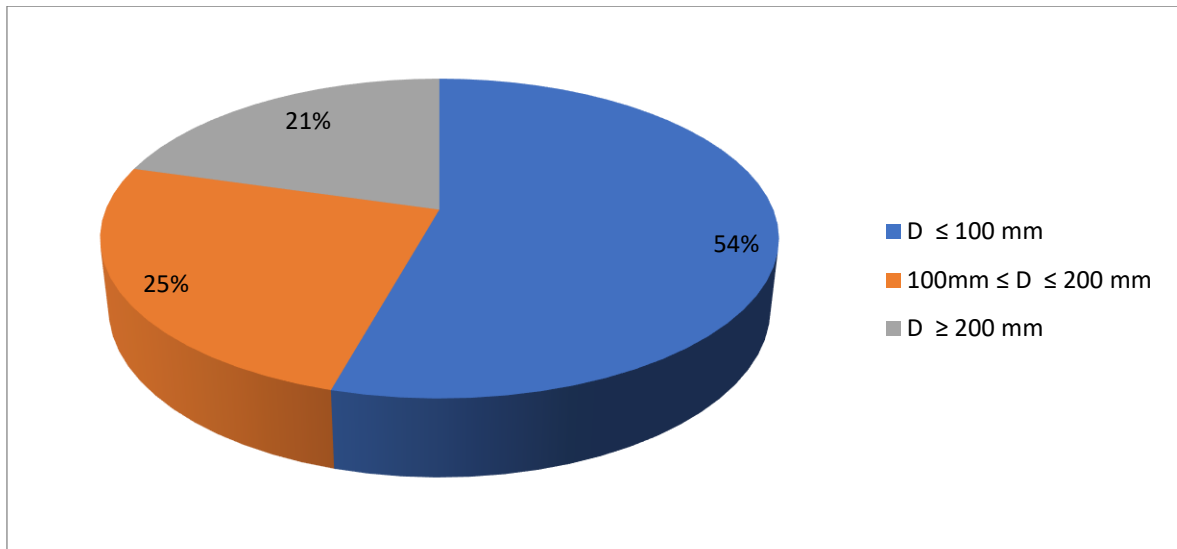


Figura 2-4 Suddivisione delle condotte in classi di diametri e rappresentazione delle percentuali

Anno di posa

Il dato relativo all'anno in cui la condotta è stata installata è presente solo per 31903 condotte su 154725, ossia il 20% del totale. La quasi totalità delle condotte che mostrano l'informazione è stata posata tra il 1990 ed il 2010, quindi solo a partire dall'ultimo trentennio si è cominciato a riportare tale dato in maniera continuativa. Considerata la notevole influenza che l'anno di posa e quindi l'età della condotta ha nel fenomeno delle rotture, si è ritenuto comunque opportuno, nonostante la mancanza di dati, prendere in considerazione laddove presente questa caratteristica. Si riporta in Tabella 2-4 la suddivisione in classi dei dati disponibili e in Figura 2-5 il corrispettivo grafico a torta.

Tabella 2-4 Analisi condotte suddivise in classi di anno di posa

Anno di posa		N°condotte	%	Chilometri	%
CLASSE 1:	1870 - 1890	114	0.4	1.3	0.3
CLASSE 2:	1890 - 1910	936	2.9	14.4	3.6
CLASSE 3:	1910 - 1930	539	1.7	10.5	2.6
CLASSE 4:	1930 - 1950	258	0.8	4.1	1.0
CLASSE 5:	1950 - 1970	456	1.4	6.5	1.6
CLASSE 6:	1970 - 1990	2692	8.4	39.3	9.8
CLASSE 7:	1990 - 2010	24968	78.3	301	75.3
CLASSE 8:	2010 →	1940	6.1	22.6	5.7
TOT		31903	100.0	399.7	100.0

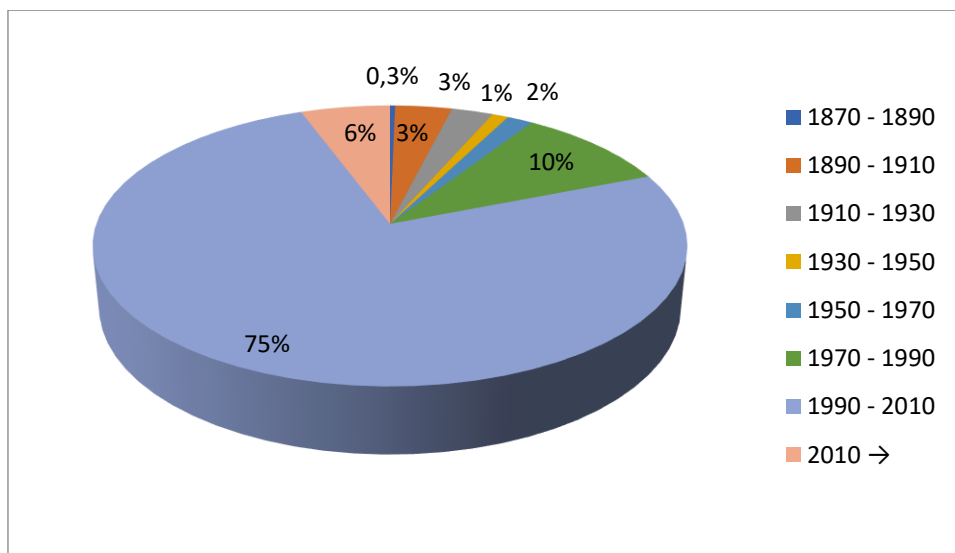


Figura 2-5 Suddivisione delle condotte in classi di anni di posa e rappresentazione delle percentuali

2.1 CONCLUSIONI

Il capitolo costituisce una presentazione della rete acquedottistica del comune di Torino a partire dai dati a disposizione del database cartografico *SMAT*.

I materiali predominanti sono la ghisa grigia e la ghisa sferoidale.

Il 68% delle tubazioni costituenti la rete sono caratterizzate da un diametro inferiore ai 100 millimetri e, infine, l'informazione sull'età, laddove presente, consente di affermare che il 70% delle condotte è stato posato nel ventennio 1990-2010.

3 LA LOGICA DI ACQUISIZIONE DEI DATI

I dati riguardanti i guasti della rete acquedottistica e fognaria dei comuni gestiti dalla Società Metropolitana Acque Torino (*SMAT*), che rappresentano la base di questo lavoro di tesi, sono stati raccolti negli anni dal personale in un unico database (la piattaforma Maximo).

Maximo è un database che facilita la gestione di tutti i tipi di asset su una singola piattaforma software. La piattaforma offre una visuale completa delle condizioni e delle collocazioni degli asset e fornisce all'azienda una capacità di pianificazione, di controllo e di verifica. Utilizzando l'interfaccia utente di Maximo Asset Management, è anche possibile monitorare le condizioni degli asset tramite dei KPI (Key Performance Indicator). Inoltre, il database fornisce un approccio modulare e, tramite la piattaforma, l'azienda memorizza tutti i dati.

Il software consente, infatti, di gestire le operazioni fin dalla segnalazione, eliminando i moduli cartacei e consentendo la conoscenza aggiornata delle segnalazioni e dello stato di avanzamento dei singoli interventi. L'operatore che esegue la riparazione si occupa anche della compilazione su tablet di un modulo in cui vengono riportate tutte le caratteristiche del guasto; tali moduli vengono salvati ed organizzati sulla piattaforma Maximo, immagazzinando l'intervento in memoria. La modalità di compilazione, però, ha lo svantaggio di lasciare libera la stesura e in molti casi l'operatore addetto alla riparazione del guasto, ha ommesso di compilare alcuni campi: affinché il modello di previsione possa disporre di un campione di dati quanto più completo possibile relativo alla rottura delle condotte, sarà necessario eseguire un notevole filtraggio dei dati.

3.1 TABELLE FONDAMENTALI

Il database contiene tutti gli interventi eseguiti sul sistema acquedottistico e fognario gestito dalla *SMAT*. I dati contenuti all'interno della piattaforma sono organizzati in tre tabelle fondamentali, fornite sotto forma di foglio elettronico. Queste tabelle contengono un numero molto elevato di informazioni e ad ogni riga della tabella è associato un intervento, mentre le colonne rappresentano i campi descrittivi della rottura.

Le tabelle fondamentali sono:

- ❖ **Workorder:** è la tabella principale e contiene i dati generali dell'intervento;
- ❖ **Woserviceaddress:** è la tabella contenente le informazioni connesse alla geolocalizzazione del guasto;
- ❖ **Vie:** è la tabella contenente i dati di indirizzo.

3.1.1 Tabella Workorder

Questa tabella è la principale alla quale verranno connesse le successive tramite il codice identificativo del guasto. Si riportano i campi che compongono la tabella *Workorder*:

- *wonum*: identifica mediante un codice numerico il guasto;
- *workorderid*: codice identificatore;
- *description*: descrizione dell'ordine di lavoro;
- *comune*: comune in cui si è verificato il guasto;
- *indirizzo*: indirizzo in cui si è verificato il guasto;
- *numeroangolo*: civico al quale si è verificato il guasto;
- *segnalazione*: codice che identifica la tipologia di segnalazione;
- *status*: descrive lo stato dell'ordine (in corso, completato);
- *reportdate*: riporta la data e l'ora dell'ordine di lavoro;
- *costiimpresa*: identifica i costi sostenuti ad un'impresa esterna per l'intervento;
- *pericolo*: indica la presenza o meno di un pericolo;
- *reparto*: indica il centro di servizio acqua;
- *datacontratto*: indica la data e l'ora della stipulazione del contratto;
- *historyflag*: cella che segnala la presenza di un record storico;
- *wopriority*: numero tra 0 e 999 indicante l'importanza dell'ordine di lavoro;
- *origrecordid*: identifica il record di origine;
- *ripristino*: indica se è stata ripristinata la condotta;
- *esitoverifica*: descrive l'esito della verifica;
- *dataverifica*: descrive la data e l'ora della verifica;
- *failurecode*: classe di guasto di primo livello;
- *problemcode*: classe di guasto di secondo livello;
- *numero mappa*: numero sopralluogo;
- *actstart*: data e ora in cui è iniziato il lavoro;
- *wol4*: data e ora di fine ripristino stradale;
- *actfinish*: data e ora in cui è stato completato il lavoro effettivo;
- *numeropresa*: numero della presa;
- *lunghezza*: lunghezze in considerazione;
- *targstartdate*: data in cui è previsto l'inizio dell'ordine di lavoro;
- *targcompdate*: data per la quale si prevede la fine dell'ordine di lavoro;
- *diametro*: diametro della presa o della condotta;
- *codicemateriale*: codice rappresentativo del materiale;
- *manovra*: descrive la presenza o meno di manovre;
- *orestop*: ore di interruzione;
- *numeromanovre*: numero di manovre effettuate;
- *preavviso*: indica se c'è stato un preavviso di 24 ore;
- *gelatura*: indica se c'è stata una gelatura della condotta;
- *actlabcost*: costo risorse;

- *acttoolcost*: costo attrezzature;
- *actmatcost*: costo materiali;
- *zanomaliapos*: indica la presenza di un'anomalia nel posizionamento dell'asset;
- *znotepos*: eventuali note sull'anomalia.

3.1.2 Tabella Woserviceaddress

Contiene i dati che identificano il punto preciso della rottura, di conseguenza ad ogni riga è associato univocamente un intervento. I campi più importanti, contenenti le coordinate geografiche, ossia latitudine e longitudine, risultano mancanti nella maggior parte dei casi per tutti gli anni, eccezion fatta per il 2016, che presenta una percentuale di copertura elevata; ciò spinge a pensare che alcuni campi siano stati aggiunti in seguito. Si riportano i campi che compongono la tabella *Woserviceaddress*:

- *wonum*: numero che identifica il guasto;
- *siteid*: identificatore della sede;
- *streetaddress*: indirizzo;
- *formattedaddress*: indirizzo formattato;
- *city*: città;
- *regiondistrict*: distretto;
- *postacode*: codice postale;
- *latitudey*: latitudine (y);
- *longitudex*: longitudine (x);
- *referencepoint*: punto di riferimento utile all'operatore per localizzare l'intervento;
- *woserviceaddressid*: codice identificatore.

3.1.3 Tabella Vie

Questa tabella contiene tutti i dati di indirizzo da associare, tramite il codice identificatore, all'intervento operato. Ogni riga non contraddistingue univocamente un guasto, poiché più ordini di lavoro fanno riferimento alla stessa riga. Si riportano i campi che compongono la tabella *Vie*:

- *comune*: comune in cui si è verificato il guasto;
- *description*: campo necessario per poter unire la tabella vie con quella principale (questo campo infatti riporta la seguente forma:comune*via/strada/piazza);
- *centrofg*: centro di servizio fognature;
- *centroh2o*: centro di servizio acqua;
- *circostrizione*;
- *indirizzo*: indirizzo in cui si è verificato il guasto;
- *vieid*: codice identificatore.

3.1.4 Estrazione dei Dati di Interesse e Calcolo delle Occorrenze

Le tabelle precedentemente descritte non contengono esclusivamente dati riferiti a rottura su condotta e, soprattutto, non tutti i dati sono utili ai fini statistici; per questo motivo sono stati estratti i soli campi di interesse per valutare la possibile relazione tra il numero di rotture e le caratteristiche delle condotte, eliminando gli elementi superflui. La selezione dei dati ha previsto di far riferimento esclusivamente alla tipologia di guasto “*fuga condotta*”, eliminare le colonne riportanti codici identificatori, descrizioni con compilazione libera e quelle riguardanti lo stato dell’intervento, ed eliminare i campi con percentuali di dati molto basse.

Per ogni tabella, estratta dal database, si riportano i soli campi di interesse e le relative occorrenze, ossia il numero di dati a disposizione per ogni campo espresso in termini percentuali.

La tabella *Workorder*, nella sua forma originale, presenta 67 colonne e 382.865 righe. Le righe identificano univocamente un guasto, mentre le colonne rappresentano i campi. I campi selezionati dalla tabella principale sono i seguenti:

- *wonum*: questo dato identifica l’intervento e, di conseguenza, è stato utile per poter unire le tre tabelle;
- *comune*: dato necessario per la localizzazione del guasto;
- *indirizzo*: dato necessario per la localizzazione del guasto;
- *reportdate*: consente di conoscere la data in cui è stata fatta la segnalazione, ma non garantisce che il guasto si sia verificato proprio in quel giorno. È utile per una collocazione temporale approssimata;
- *costimpresa*: questo è il costo sostenuto dalla SMAT in seguito all’intervento di un’impresa esterna. Se il guasto presenta in questo record uno zero significa che la riparazione è stata eseguita da uno o più operatori SMAT;
- *pericolo*: dato preso in considerazione per poter valutare la presenza di un pericolo;
- *diametro*: caratteristica della condotta;
- *codicemateriale*: caratteristica della condotta;
- *orestop*: dato necessario per valutare il numero di ore di interruzione del servizio;
- *problemcode*: la classe di guasto indicata da questa colonna è stata fondamentale per identificare soltanto gli interventi riguardanti le fughe su condotta.

Tabella 3-1 Occorrenze dei campi di interesse della tabella *Workorder*

Workorder					
Campo	Wonum	Comune	Indirizzo	Reportdate	Costiimpresa
%	100	100	100	99.9	99.9
Campo	Pericolo	Diametro	Codicemateriale	Orestop	Problemcode
%	100	35.4	26.5	9.4	83

La tabella *Vie* di origine presenta 9 colonne e 22.779 righe e, come già esposto, le righe di questa tabella non individuano univocamente un guasto, per questo motivo il numero di righe non coincide con quelle della tabella *Workorder*. Vengono di seguito riportati i campi presi in considerazione:

- *comune*: dato necessario per la localizzazione del guasto;
- *indirizzo*: dato necessario per la localizzazione del guasto;
- *description*: dato necessario per poter unire questa tabella con la tabella *Workorder*;
- *circostrizione*: dato necessario per la localizzazione del guasto.

Tabella 3-2 Occorrenze dei campi di interesse della tabella *Vie*

Vie				
Campo	Comune	Indirizzo	Description	Circoscrizione
%	100	99.9	100	98.6

Infine, la tabella *Woserviceaddress* presenta lo stesso numero di righe della tabella *Workorder* e 32 colonne. Di questa si sono considerati i seguenti campi:

- *wonum*: questo dato è necessario perché identifica l'intervento ed è stato utilizzato per l'unione delle tabelle;
- *latitudey*: dato necessario per poter geolocalizzare il guasto;
- *longitudey*: dato necessario per poter geolocalizzare il guasto.

Tabella 3-3 Occorrenze dei campi di interesse della tabella *Woserviceaddress*

Woserviceaddress			
Campo	Wonum	Latitudey	Longitudey
%	100	17	17

Dall'analisi delle Tabella 3-1, Tabella 3-2 e Tabella 3-3, si evince che il database, contenente le segnalazioni pervenute tra l'anno 2006 e l'anno 2016, permette di identificare l'intervento tramite il codice identificatore *wonum* e i campi *indirizzo* e *comune* nel 100% dei casi. Molti campi, soprattutto *diametro*, *codicemateriale*, *latitudey* e *longitudey*, sono invece privi di informazioni complete.

Selezionati i campi di interesse, utili ai fini di una statistica, è possibile unire le tre tabelle attraverso dei record identificatori contenuti all'interno di ciascuna tabella: il campo *Wonum* per le tabelle *Workorder* e *Woserviceaddress*, ed il campo *Indirizzo* per la tabella *Vie*, ottenendo un'unica tabella contenente le informazioni che definiscono gli interventi.

Si riporta in Figura 3-1 il flow chart delle operazioni eseguite per ottenere l'unione delle tre tabelle.

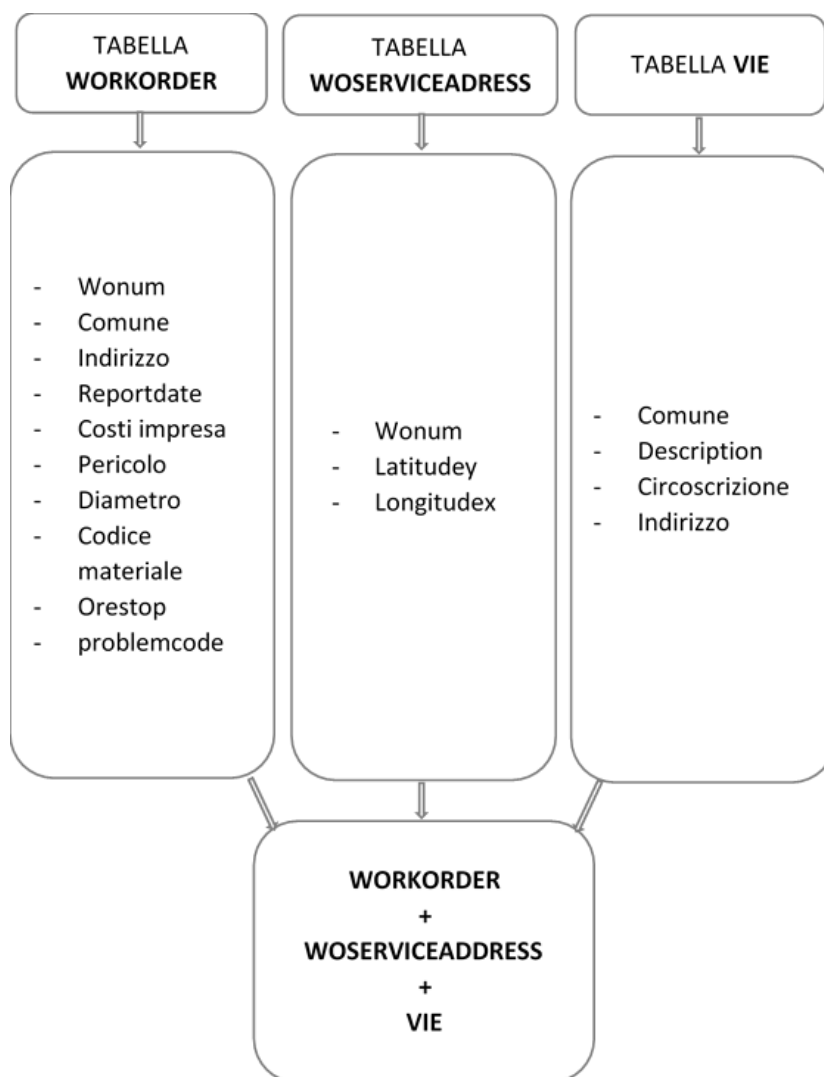


Figura 3-1 Estrazione dei dati e unione delle tre tabelle

3.2 TABELLA WwvCONDOTTETORINO

La tabella ottenuta dall'unione delle tre fondamentali, contenente gli interventi di tipo “*fuga condotta*” segnalati ed eseguiti nel comune di Torino dall'anno 2006 all'anno 2016, presenta percentuali eterogenee di dati in ogni campo di interesse.

Tabella 3-4 Occorrenze dei campi della tabella contenente gli interventi “Fuga Condotta” nel comune di Torino

Workorder+Vie+Woserviceaddress						
Campo	Wonum	Comune	Indirizzo	Reportdate	Costiimpresa	Pericolo
%	100	100	100	100	100	100
Campo	Diametro	Codicemateriale	Orestop	Circoscrizione	Latitudey	Longitudex
%	90.7	87.8	76.5	99.8	12	12

La Tabella 3-4 mostra la presenza di un'elevata percentuale di dati nella maggior parte dei campi, eccezion fatta per i campi *latitudey* e *longitudex*, che presentano solo il 12% di dati presenti. Questi ultimi due sono di fondamentale importanza per localizzare geograficamente in modo preciso il guasto e, in seguito, calcolare, attraverso un modello idraulico della rete sviluppato in ambiente *Epanet*, la pressione della condotta dove è stato eseguito l'intervento di riparazione.

Al fine di ottenere informazioni riguardanti la geolocalizzazione del guasto, è stato scritto un codice in *Python* che, attraverso una chiave di *Google Maps*, ha restituito latitudine e longitudine degli interventi dotati di indirizzo completo di numero civico. Effettuata la georeferenziazione, la percentuale di occorrenze è aumentata al 95% ed è stato quindi possibile riportare le rotture sulla rete di distribuzione di Torino tramite l'applicazione *QGIS Desktop*, come mostra la Figura 3-2.

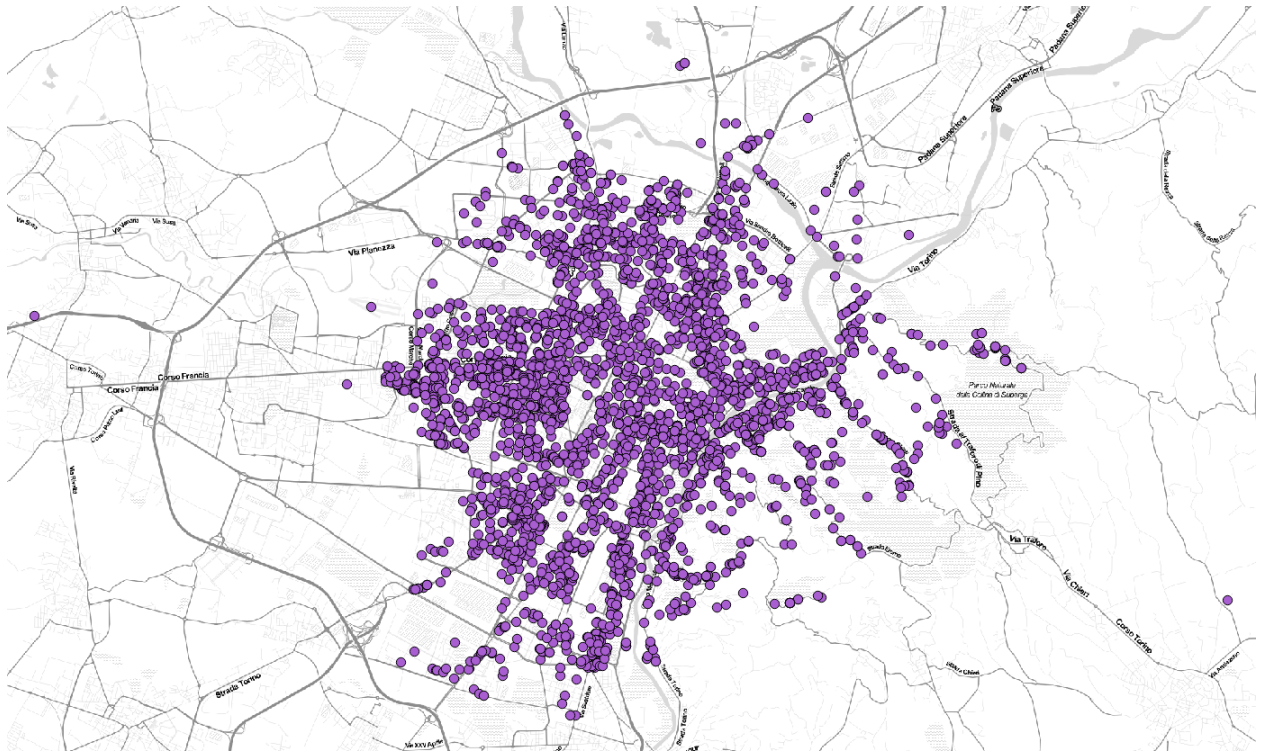


Figura 3-2 Distribuzione delle rotture di tipo “Fuga Condotta” all’interno della rete del comune di Torino

Integrando la tabella ottenuta dall’unione delle tre principali delle informazioni relative alla pressione, utilizzando il modello idraulico della rete, e dell’informazione relativa all’anno di posa delle condotte è stata ottenuta la tabella *Wwvcondottetorino*, che costituisce il punto di partenza per la pre-analisi dei dati utili ai fini statistici.

Questa tabella, nella sua versione finale, presenta gli interventi di tipo “fuga condotta” che hanno avuto luogo nel comune di Torino tra il 2006 e il 2016 ed è caratterizzata da 4042 righe, ossia i guasti, e 14 colonne, che rappresentano i campi contenenti le informazioni di ogni guasto.

Si riporta di seguito un esempio di guasto descritto nella tabella *Wwvcondottetorino*.

Tabella 3-5 Esempio di guasto nella tabella *Wwvcondottetorino*

Tabella Wwvcondottetorino						
Wonum	Comune	Indirizzo	Pericolo	Reportdate.1	Diametro	Codicemateriale
850347	Torino	Viale Stefano Turr	0	17/12/2016 15:04	100	Ghisa Grigia
Orestop	Costiimpresa.1	Circoscrizione	Latitudey	Longitudex	Pressione	Anno di posa
2	731.62	8	45.0463632	7.684207	68.105	1966

La Figura 3-3 mostra il flow chart delle operazioni eseguite per ottenere la tabella *Wwvcondottetorino*.

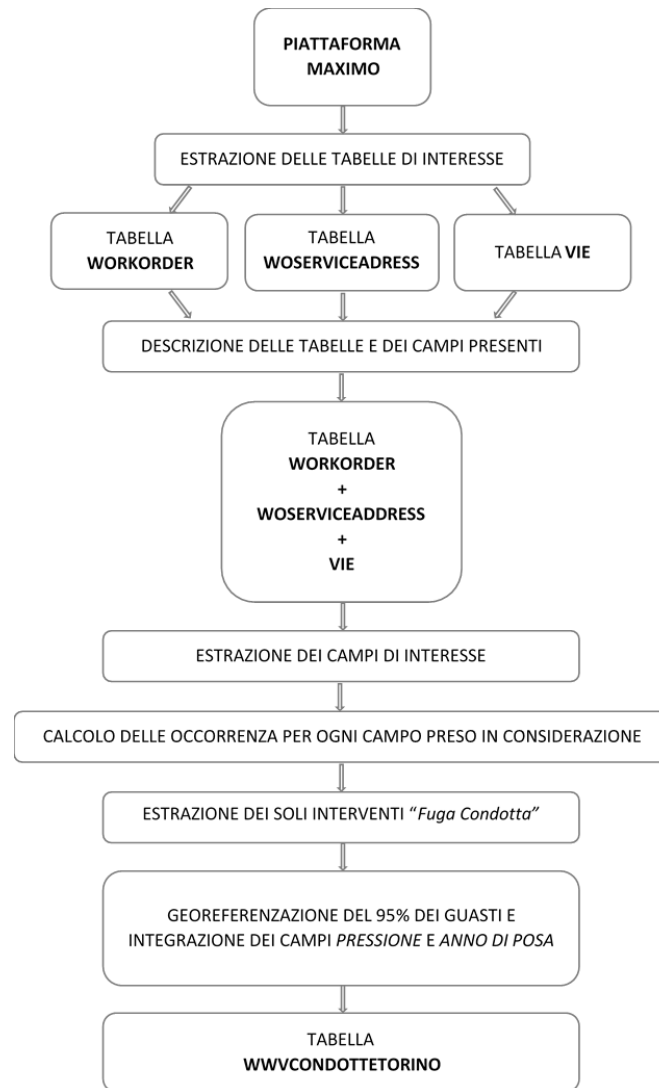


Figura 3-3 Procedura eseguita per ottenere la tabella Wwvcondottetorino

3.3 TABELLA RETE TORINO

Come accennato nei capitoli precedenti, il fine di questo studio è l'applicazione di un modello statistico di regressione ai dati a disposizione per poter prevedere quali condotte siano a rischio di rottura. Affinché ciò avvenga, è necessario che i modelli siano tarati a partire da un'unica tabella contenente tutte le informazioni sulle condotte, compreso il dato di rottura o non rottura, che costituirà la variabile dipendente dicotomica Y del modello.

Sebbene lo sviluppo del modello di regressione verrà trattato nei capitoli successivi, questa breve introduzione è necessaria al fine di chiarire il motivo per cui si è scelto di creare un unico quadro rappresentativo di tutte le condotte che compongono il sistema acquedottistico.

La regressione logistica prevede di costruire un modello in cui la variabile dipendente Y presenti solo due possibili risultati (0 e 1), che descrivono l'esito, cioè il successo o l'insuccesso del verificarsi di un evento aleatorio. In questo ambito applicativo, i due possibili risultati che la variabile Y può assumere sono:

- la condotta si è rotta (1);
- la condotta non si è rotta (0).

Le variabili esplicative x possono essere, invece, sia di tipo qualitativo che quantitativo e, in questo caso, sono:

- materiale;
- diametro;
- carico massimo;
- anno di posa;
- variazione di carico.

Solo in seguito verranno studiate ed inserite nel modello le variabili dovute ai fattori ambientali, quali i carichi dinamici:

- velocità media;
- flusso veicolare;
- tram route.

I passaggi seguiti per ottenere la tabella denominata *ReteTorino*, contenente tutte le caratteristiche delle condotte, sono i seguenti:

- è stato utilizzato il database cartografico unitamente al modello della rete elaborato in ambiente *Epanet*, in grado di fornire i valori di pressione che caratterizzano le condotte della rete di Torino. È stato così possibile conoscere la totale composizione della rete, le caratteristiche strutturali, geometriche e i carichi nelle condotte;
- sono state escluse dall'analisi le condotte con lunghezza inferiore ai 10 metri, in quanto insufficientemente significative ai fini della stima del modello di regressione. Conseguenza di tale scelta, è stata l'eliminazione di una notevole quantità di condotte in PEAD, poiché caratterizzate da lunghezze molto contenute;
- alle condotte presenti nel database cartografico sono state associate attraverso le coordinate geografiche quelle che costituiscono la tabella *Wwvcondottetorino* (che contiene esclusivamente informazioni sulle condotte che hanno presentato rottura nel periodo di riferimento 2006-2016). Va sottolineato che solo il 95% degli elementi contenuti nella tabella *Wwvcondottetorino* dispongono delle informazioni geografiche ed in alcuni casi le informazioni non coincidono con quelle presenti nel database cartografico, probabilmente per errore di trascrizione da parte

dell'operatore che ha eseguito l'intervento o per errore di geolocalizzazione in funzione dell'indirizzo del complesso più vicino alla rottura. Ciò non ha consentito l'associazione del guasto ad alcune condotte. Dopo aver inserito tutti gli interventi contenuti nella tabella *Wwvcondottetorino* sulla cartografia della rete, questi sono stati associati alla condotta più vicina, nel raggio di 500 metri, con le medesime caratteristiche (materiale, diametro);

- nel caso di coincidenza tra le informazioni geografiche riportate in tabella e il database cartografico, è stata assegnata l'informazione di rottura, contraddistinta da flag pari a 1. Ai restanti elementi è stato assegnato flag pari a 0, indice dell'assenza di guasti;
- qualora la medesima condotta avesse presentato più rotture nell'intervallo di tempo considerato, è stato riportato nell'apposito campo il numero totale di rotture.

La tabella *ReteTorino*, così ottenuta, contiene le condotte facenti parte della rete di Torino e le informazioni riguardo la presenza o assenza di rottura nel periodo di osservazione. Questa è composta da:

- 25.718 righe, ognuna delle quali fa riferimento, in maniera univoca, ad una condotta;
- 12 colonne, che rappresentano gli elementi utili per il calcolo del tasso di rottura e per l'applicazione della regressione logistica.

Si riporta a titolo esemplificativo un elemento riga della tabella *ReteTorino*.

Tabella 3-6 Esempio di caratteristiche della condotta appartenente alla tabella ReteTorino

Tabella <i>ReteTorino</i>					
Id_pipe	Materiale	Anno_Posa	Latitudine	Longitudine	Diametro [mm]
8720654	GhSf	1993	49.86965933	13.98612484	100
Pmax [m]	Prange [m]	Pmean [m]	Lunghezza [m]	Flag	N° Rottura
3.4492	1.8315	2.4245	307	1	1

Delle condotte presenti in tabella, 2870 presentano flag pari a 1, cioè si sono rotte nell'arco temporale compreso tra il 2006 e il 2016; la percentuale di condotte rotte rispetto al numero totale di condotte presenti in tabella è pari a 11.16% ($2870 \cdot 100 / 25718$).

Poiché alcune condotte hanno presentato più volte rottura, come indicato nel campo '*N Rottura*', il totale dei guasti è di 3881 e la percentuale di rotture, nell'arco di tempo indicato, in funzione del numero totale di condotte presenti all'interno della rete è del 15.10% ($3881 \cdot 100 / 25718$).



Figura 3-4 Distribuzione delle rotture e delle condotte costituenti la rete del comune di Torino

Nonostante la tabella ottenuta rappresenti la rete di Torino, la sua struttura non risulta adeguata per l'applicazione del modello di regressione; questo necessita della totale copertura dei campi contenenti le informazioni per la stima dei parametri, mentre la tabella *ReteTorino* è così caratterizzata:

- tutte le condotte dispongono dell'informazione sul materiale;
- tutte le condotte dispongono dell'informazione sul diametro;
- solo 6033 condotte su 25718 dispongono dell'informazione sull'anno di posa. La copertura di questo campo è del 23.46%;
- 24660 condotte su 25718 dispongono dell'informazione sul carico massimo. La copertura di questo campo è del 95.88%. Alcuni dati di carico sono però certamente sbagliati, in quanto presentano valore negativo;
- solo 24660 condotte su 25718 dispongono dell'informazione sulla variazione di carico. La copertura di questo campo è del 95.88%. Alcuni dati di carico sono però certamente sbagliati, in quanto presentano valore negativo.

Il modello di regressione deve fare affidamento su dati completi, privi di errori e con totale copertura dei campi, per tale ragione, è stata elaborata una nuova tabella descritta nel paragrafo successivo.

3.4 TABELLA REGRESSIONECONDOTTE

Sulla tabella *RegressioneCondotte* verrà applicato il modello di regressione logistica, in cui le variabili esplicative saranno le caratteristiche delle condotte, ossia i fattori strutturali (anno di posa, diametro e materiale) e su fattori operativi (pressione massima di esercizio e variazione di carico). La realizzazione di questa ha previsto l'estrazione di sole quelle condotte che possiedono tutte e cinque le variabili indipendenti x , ossia i fattori elencati tra parentesi, dalla tabella *ReteTorino*.

Inoltre i materiali presenti in tabella, che rappresentano una variabile qualitativa e non quantitativa, sono stati ordinati in funzione del valore decrescente del tasso di fallanza ed è stato assegnato loro un numero intero crescente, che sarà introdotto nel modello per la stima dei parametri.

Tabella 3-7 Codifica numerica dei materiali presenti in rete

Materiale	Tasso di fallanza [NR/(km*anno)]	Num. materiale
Eternit	5.94	1
Ghisa grigia	4.03	2
Ferro	2.42	3
Ghisa sferoidale	0.97	4
Chameroy	0.85	5
PEAD	0.71	6
Acciaio	0.70	7
PVC	0.61	8
Cemento armato	0.34	9
Materiali indefiniti	0.00	10

Poiché durante l'analisi della tabella *ReteTorino* sono stati riscontrati dei problemi di non coincidenza tra i dati di questa e quelli contenuti nella tabella delle rotture *Wwvcondottetorino*, causati dalla errata geolocalizzazione dei guasti, sono stati eliminati tutti gli elementi aventi caratteristiche non coincidenti con quelle delle condotte associate.

La rimozione di tutte le condotte prive di uno o più variabili indipendenti, l'assegnazione di un numero al campo materiale e la rimozione delle condotte per cui non si ha coincidenza tra le tabelle *ReteTorino* e *Wwvcondottetorino*, conducono alla realizzazione della tabella *RegressioneCondotte*, che è composta da:

- 5740 elementi, che rappresentano le condotte filtrate con le modalità sopra descritte;
- 87 condotte su 5740 che presentano flag pari a 1. La percentuale di condotte rotte è del 1.52% ($87 \cdot 100 / 5740$);
- alcune condotte hanno presentato più rotture nell'intervallo di tempo di riferimento, per un totale di 106 rotture su 5740 condotte ed una percentuale di rottura di 1.85%.

Tutti i 5740 elementi della tabella *RegressioneCondotte* sono caratterizzate dalla totale disponibilità di informazioni riguardanti materiale, diametro, anno di posa, carico massimo, variazione di carico e presenza\assenza rottura.

L'eliminazione delle condotte che non presentavano coincidenza di caratteristiche o che presentavano campi incompleti, ha drasticamente ridotto la percentuale di rotture, che passa dal 15.10% della tabella *ReteTorino* al solo 1.85% della tabella *RegressioneCondotte*.

3.5 TABELLA REGRESSIONE TGM

A partire dalla tabella *RegressioneCondotte* è stata elaborata la tabella *RegressioneTGM*, che presenta, oltre a quelli già descritti, campi relativi ai fattori ambientali, in particolare al carico dinamico, impersonato dai dati di velocità media e traffico giornaliero medio (*TGM*). Per ottenere la tabella nella sua forma finale, verranno eseguiti dei passaggi che escluderanno parecchie condotte, ma queste operazioni saranno necessarie per poter basare la regressione esclusivamente su tratti aventi tutte le informazioni utili; la tabella sarà, infatti, utilizzata per la stima del modello di regressione e la susseguente fase di testing.

I dati di traffico sono stati forniti dalla società *5T Srl*¹², che opera nel campo dei sistemi ITS (IntelligentTransport Systems) e dell'infomobilità.

I dati, forniti su foglio elettronico denominato *TGM_2016*, forniscono i valori del *TGM* e delle velocità medie stimate per l'anno 2016 sulle strade presenti nel modello del Supervisore Metropolitano di *5T*. I vari archi stradali sono caratterizzati in tabella da un codice, diverso nei due sensi di marcia, e dalle coordinate degli estremi, espresse nel sistema di riferimento *WGS84*.

A partire dai valori di *TGM* è possibile ottenere i dati di flusso veicolare, espressi in veicoli\ora, dividendo il dato di partenza per 24.

Utilizzando i dati sul servizio *GTT (Gruppo Torinese Trasporti)*, forniti dalla società stessa, in formato GTFS (General Transit Feed Specification), che definisce un format comune per gli orari dei trasporti pubblici e le relative informazioni geografiche, è stato aggiunto un campo al foglio elettronico *TGM_2016* fornito da *5T*, per indicare la presenza di binari dedicati al passaggio dei tram su ogni arco stradale.

La scelta di sottolineare la presenza del mezzo è dovuta alle vibrazioni che esso produce, che potrebbero generare stress alle condotte. Il campo aggiunto, denominato *TramRoute*, assume valore pari a:

¹²*5T Srl* è una società a totale partecipazione pubblica. Le quote azionarie di *5T* sono detenute da: *GTT Spa*, Città di Torino, Provincia di Torino e Regione Piemonte.

- 1 se l'arco stradale è sede della linea tranviaria;
- 0 se l'arco stradale non è soggetto al passaggio del tram.

La Tabella 3-8 mostra le prime due righe della tabella *TGM_2016*, a cui è stato aggiunto il campo *TramRoute*: la prima riga contiene i campi considerati, mentre la seconda riporta le informazioni di un arco stradale.

Tabella 3-8 Esempio informazioni arco stradale

TGM_2016								
Id_arc	Lat_A	Lng_A	Lat_B	Lng_B	Flussi [veic/h]	TGM	Velocità [m/s]	Tram Route
588-5016-1	45.09743	7.633898	45.09364	7.632451	1094	26256.24	38.514	0

La tabella *RegressioneCondotte* e quella contenente i dati di traffico *TGM_2016* non hanno campi in comune, che potrebbero facilitarne l'unione. Questa è avvenuta attraverso l'analisi spaziale utilizzando l'applicazione *QGIS Desktop* e i passaggi seguiti per ottenere la tabella *RegressioneTGM* sono i seguenti:

- sono stati riportati su mappa gli elementi delle tabelle *RegressioneCondotte* e *TGM_2016*. Da questo primo passaggio risulta chiaro che molti elementi condotta verranno eliminati dalla tabella poiché sono molte le strade, soprattutto secondarie, di cui non si hanno dati di traffico a disposizione, ma che sono interessate dal passaggio delle tubazioni. Si riporta un estratto di mappa in Figura 3-5 in cui la problematica risulta chiara;



Figura 3-5 Estratto di mappa in cui sono riportati i dati della tabella TGM_2016 (in nero) e della tabella RegressionCondotte (in verde)

- al fine di considerare la reale larghezza stradale e non limitarsi allo spessore millimetrico della linea rappresentante il solo asse stradale, è stata creata una zona buffer attorno ai vettori polilinee costituiti dagli archi stradali della tabella TGM_2016. Il buffering crea un'area che è entro una specificata distanza su entrambi i lati del vettore per le geometrie selezionate, in questo caso di 20 metri. Tale valore è stato scelto considerando la larghezza della carreggiata, composta da una o più corsie di marcia, e l'eventuale presenza di controviali;



Figura 3-6 Zona buffer degli archi stradali presenti in tabella TGM_2016

- tramite gli strumenti di geoprocessing è stata effettuata l'intersezione spaziale tra le zone di buffer e la tabella *RegressioneCondotte*; l'intersezione spaziale utilizza le informazioni spaziali per estrarre nuovi e ulteriori significati dai dati GIS. La nuova tabella ottenuta presenta 7451 righe, che fanno riferimento alle condotte, molte delle quali più volte presenti. La presenza di doppioni è dovuta a zone buffer sovrapposte nel caso di archi stradali a doppio senso di marcia e in corrispondenza degli incroci. Qualora una condotta si trovasse in corrispondenza della sovrapposizione delle zone buffer, l'intersezione spaziale genera più righe della stessa condotta, riportando le caratteristiche della condotta stessa, uguali per ogni riga, e aggiungendo nei campi relativi ai dati di traffico i valori di ogni zona buffer sovrapposta.

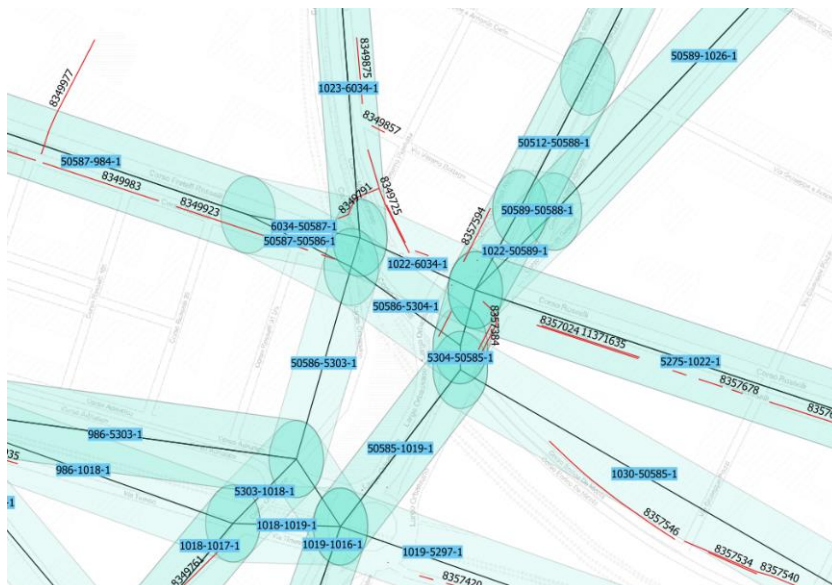


Figura 3-7 Intersezione spaziale tra zone buffer e la tabella *RegressioneCondotte*

Si riportano, a titolo d'esempio, le informazioni di una condotta, posta in corrispondenza di un incrocio, associata a più archi stradali, che sono caratterizzati da diversi valori di traffico.

Tabella 3-9 Condotta associata a più archi stradali in seguito all'intersezione spaziale

idPipe	AnnoPosa	diametro	pMax	flag	lunghezza	N Rottura	id_arc	flusso	TGM	velocità	Tram Route
25411992	1983	200	38.2205	0	139	0	9214-9213-1	305.0845	7322.028	23.850	0
25411992	1983	200	38.2205	0	139	0	9213-9214-1	332.5166	7980.3984	25.158	0
25411992	1983	200	38.2205	0	139	0	9213-9227-1	202.7282	4865.4768	25.497	0
25411992	1983	200	38.2205	0	139	0	5311-9213-1	301.0985	7226.364	23.041	0
25411992	1983	200	38.2205	0	139	0	9213-5311-1	359.4392	8626.5408	26.052	0
25411992	1983	200	38.2205	0	139	0	9227-9213-1	274.6505	6591.612	25.090	0

- al fine di analizzare una sola volta le condotte, alcune delle quali evidentemente sollecitate da più strade confluenti o adiacenti, si è proceduto in maniera diversa in funzione dei valori numerici del flusso veicolare e ancor meglio in funzione della differenza tra il valore massimo e minimo, in particolare:
 - se la differenza tra il massimo e il minimo del flusso è minore di 200 veic/h, è stato assegnato alla condotta il valore di flusso veicolare e velocità media ottenuto effettuando la media aritmetica di tutti i valori caratterizzanti la stessa condotta;
 - se la differenza tra il massimo e il minimo del flusso è maggiore di 200veic/h, ad ogni condotta è stato assegnato manualmente il valore di flusso e velocità media dell'arco stradale sovrastante.

Si riporta in Tabella 3-10 e in Tabella 3-11 il procedimento attuato nei due casi.

Tabella 3-10 Differenza tra il massimo e il minimo del flusso minore di 200veic/h

idPipe	flussi [veic/h]	Procedimento attuato
6866988	144.16	Min flussi 58.70
6866988	58.70	Max flussi 190.33
6866988	184.60	Max-Min flussi 131.63
6866988	190.34	Media flussi 130.38
6866988	88.18	Flussi [veic/h] 130.38
6866988	116.35	

Tabella 3-11 Differenza tra il massimo e il minimo del flussomaggiore di 200veic/h

idPipe	flussi [veic/h]	Procedimento attuato
6867500	2264.35	Min flussi 58.70
6867500	58.70	Max flussi 3747.96
6867500	190.34	Max-Min flussi 3689.25
6867500	3747.96	Media flussi -
		Flussi [veic/h] Associato manualmente

La tabella *RegressioneTGM*, ottenuta in seguito ai passaggi descritti, è composta da:

- 2548 elementi, che rappresentano le condotte;
- 49 condotte su 2548 che presentano flag pari a 1. La percentuale di condotte rotte è del 1.92% ($49 \cdot 100 / 2548$);
- alcune condotte hanno presentato più rotture nell'intervallo di tempo di riferimento, per un totale di 63 rotture su 2548 condotte ed una percentuale di rottura di 2.47%

Tutti i 2548 elementi della tabella *RegressioneTGM* sono caratterizzate dalla totale disponibilità di informazioni riguardanti materiale, diametro, anno di posa, carico massimo, variazione di carico, presenza\assenza rottura, velocità media, flusso veicolare e presenza\assenza passaggio tram.

L'eliminazione delle condotte poste in corrispondenza di strade di cui non si hanno informazioni ha ulteriormente ridotto il numero di elementi da analizzare. La percentuale di rotture passa dal 15.10% della tabella *ReteTorino* al 2.47% della tabella *RegressioneTGM*.

3.6 CONCLUSIONI

Questo capitolo descrive le fonti dei dati dalle quali è stato attinto. Dalla piattaforma *Maximo*, utilizzata da *SMAT*, sono state estratte le tre tabelle fondamentali *Vie*, *Workorder* e *Woserviceaddress*. Di queste sono stati analizzati soprattutto i campi di interesse e gli interventi '*Fuga Torino*' che hanno avuto luogo nel comune di Torino tra il 2006 e il 2016, generando un'unica tabella, *Wwvcondottetorino*. Attraverso un codice in *Python* è stato possibile aggiungere le informazioni di latitudine e longitudine, fino a raggiungere una copertura del 95% totale dei guasti in questi campi. Utilizzando la georeferenziazione, il database cartografico e il modello idraulico in *Epanet*, sono state aggiunte le informazioni relative agli anni di posa e alle pressioni delle condotte che hanno riscontrato dei guasti. In seguito, attraverso la geolocalizzazione è stata aggiunta alle condotte del database cartografico l'informazione riguardante la presenza di rottura o non rottura avvenuta nell'intervallo di tempo 2006-2016. L'intersezione tra la tabella *Wwvcondottetorino* e il database cartografico ha dato così luce alla tabella *ReteTorino*.

La tabella *ReteTorino* presenta molte informazioni mancanti e alcuni errori di localizzazione delle condotte; per poter ottenere una tabella da utilizzare per l'applicazione del modello di regressione, sono state prese in considerazione solo le condotte caratterizzate da totale copertura dei dati per i campi: materiale, anno di posa, diametro, carico massimo, variazione di carico e flag. Inoltre, sono stati eliminati i guasti caratterizzati da materiale e diametro non coincidenti con le caratteristiche riportate nella tabella *Wwvcondottetorino*. La tabella ottenuta è stata denominata tabella *RegressioneCondotte*.

Infine è stata effettuata l'intersezione spaziale tra la tabella *RegressioneCondotte* e i dati di traffico forniti dalla società *5T*, nonché i dati sul servizio *GTT* in merito ai passaggi dei tram, ottenendo la tabella *RegressioneTGM* caratterizzata da totale copertura dei campi: materiale, anno di posa, diametro, carico massimo, variazione di carico, velocità media, flusso veicolare, presenza/assenza passaggio tram e flag.

Nel capitolo successivo verranno analizzate le tabelle ricavate e sarà presentata la definizione di tasso di fallanza, utile nel definire la vulnerabilità di classi di condotte.

Si riporta, infine, il flowchart delle operazioni effettuate che hanno condotto alla creazione della tabella *RegressioneTGM*.

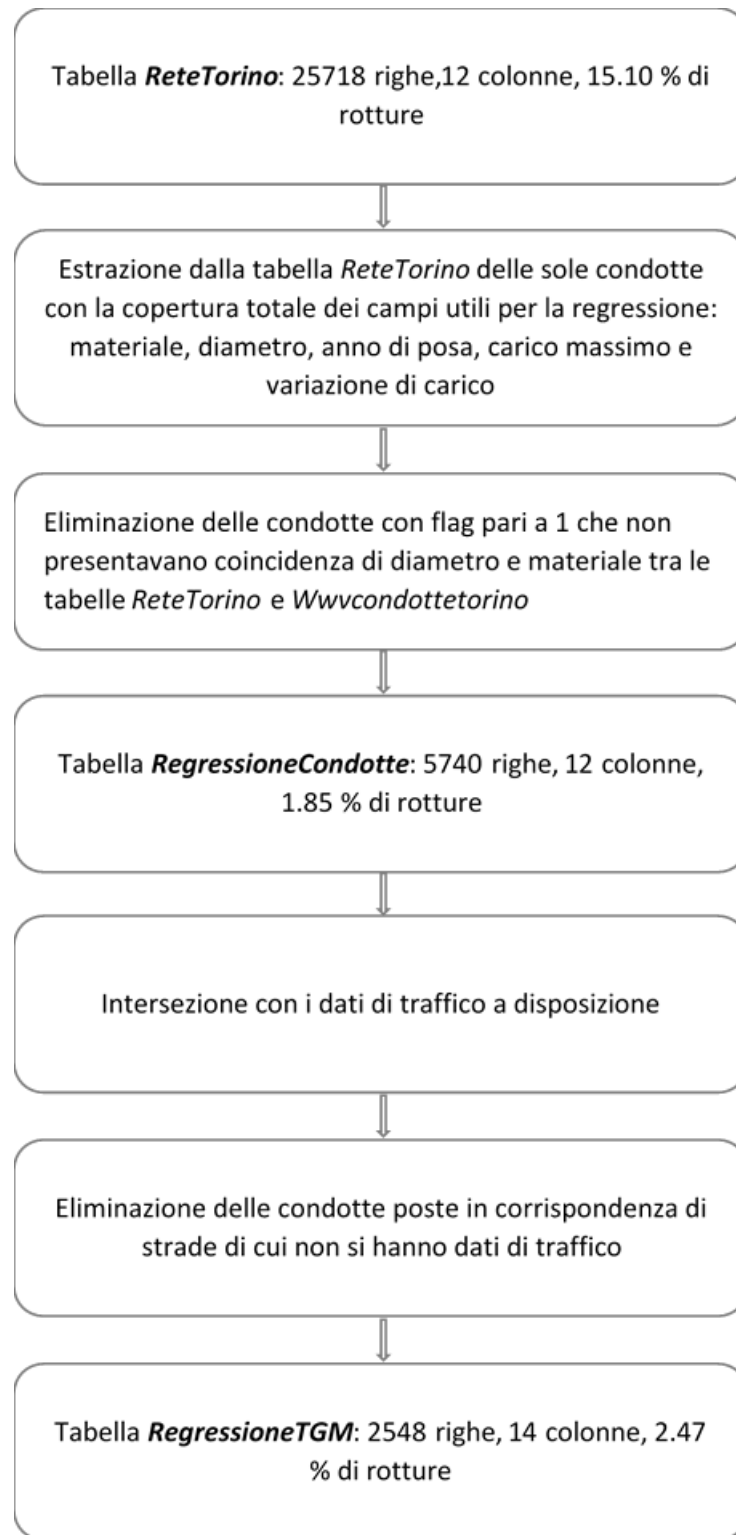


Figura 3-6 Diagramma di flusso delle operazioni eseguite per ottenere la tabella *RegressioneTGM*

4 ANALISI DEI DATI IN FUNZIONE DI FATTORI STRUTTURALI, OPERATIVI E AMBIENTALI

Nei capitoli precedenti si sono riportate le fonti dei dati e la modalità di costruzione delle tabelle. Nel seguente capitolo verranno presentate le caratteristiche e le peculiarità dei dati, che rappresentano le variabili d'interesse, raccolti nelle tabelle *ReteTorino*, *RegressioneCondotte*, e *RegressioneTGM*. Lo studio dei dati consentirà di effettuare un'analisi di vulnerabilità della rete di distribuzione idrica di Torino, motivo per il quale è opportuno introdurre la definizione del *tasso di fallanza* o *tasso di rottura*. Nello specifico questo è un indicatore che permette di valutare le prestazioni delle condotte idriche e che permette lo studio di misure di miglioramento e la programmazione ottimale degli interventi da realizzare. L'equazione che consente di calcolare la grandezza appena introdotta può essere scritta come:

$$\lambda = \frac{NR}{L * T} \quad (1)$$

Dove è indicato con *NR* il numero di rotture totali delle condotte appartenenti ad una determinata classe, *L* è la lunghezza in chilometri delle condotte appartenenti alla classe scelta e *T* è il periodo di osservazione. In questo studio il calcolo del tasso di rottura è stato eseguito suddividendo in classi i materiali, i diametri, gli anni di posa, i carichi massimi, le variazioni di carico, le velocità medie, il flusso veicolare e la presenza del tram.

Il procedimento attuato si articola nei seguenti step:

- sono state suddivise in classi le caratteristiche, appena citate, delle condotte;
- per ogni classe è stata calcolata la lunghezza complessiva delle condotte e il numero di rotture che le appartengono (condotte con flag pari a 1);
- è stata applicata la relazione (1), ricavando il tasso di fallanza per ogni classe.

Le caratteristiche analizzate per ogni tabella e i passaggi effettuati sono richiamati nello schema in Figura 4-1.

Si riportano quindi le occorrenze calcolate nelle tre tabelle per tutte le classi. È necessario sottolineare che i risultati sono influenzati dal filtraggio, dovuto ad una non perfetta geolocalizzazione delle condotte e degli interventi di ripristino della rete in seguito alla segnalazione del guasto.

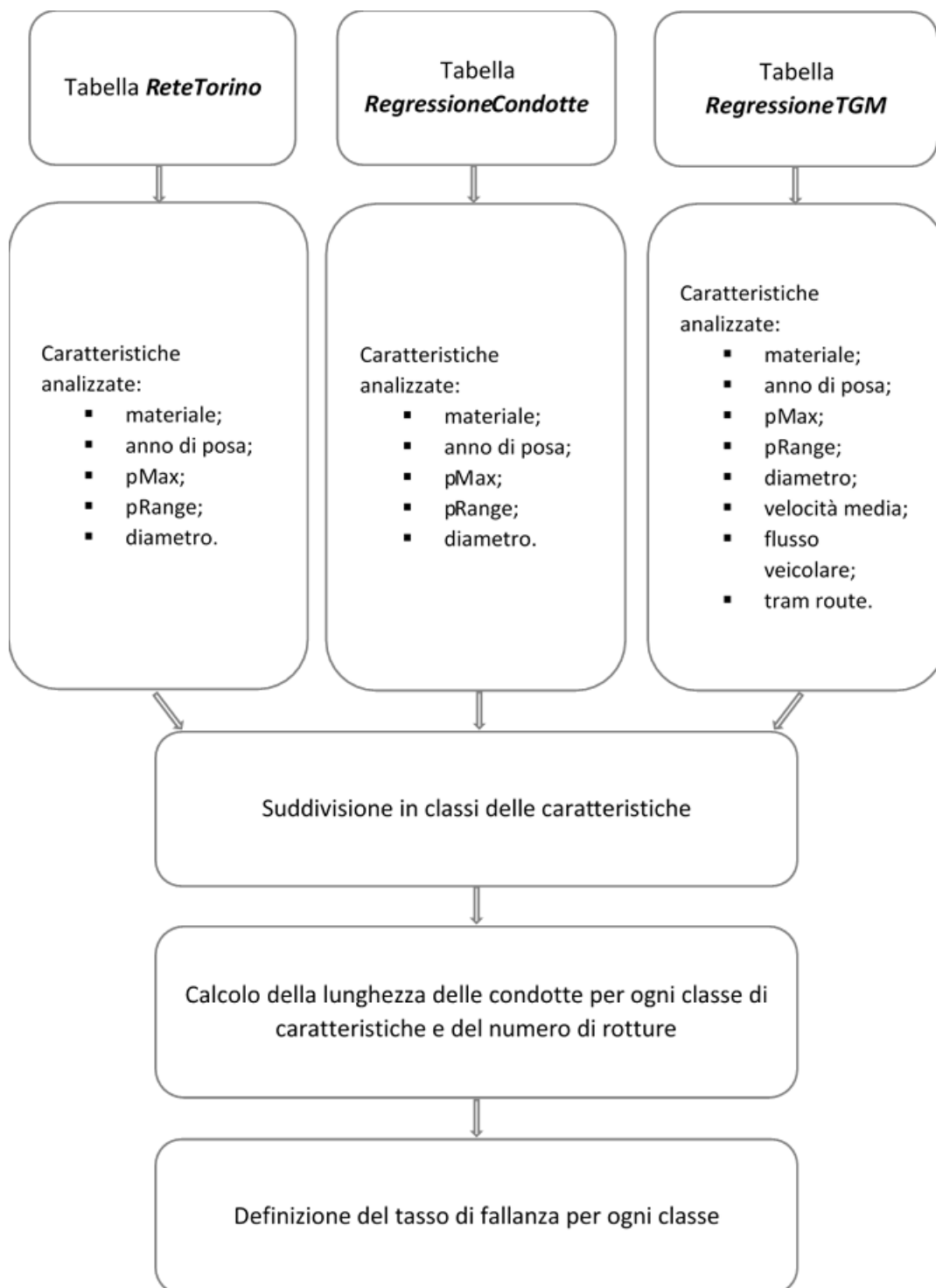


Figura 4-1 Diagramma di flusso delle operazioni eseguite per ottenere il tasso di fallanza

4.1 ANALISI DELLA TABELLA RETE TORINO

L'analisi della tabella *ReteTorino* è avvenuta suddividendo in classi i fattori strutturali e operativi: materiale, diametro, anno di posa, carico massimo e variazione di carico. Lo stesso procedimento è stato applicato alle sotto-tabelle estratte da *ReteTorino* relative ad ogni materiale presente in rete.

Calcolo del tasso di fallanza delle classi di materiale

Le condotte della tabella *ReteTorino* sono state suddivise in classi di materiale e sono state calcolate le lunghezze e le occorrenze delle rotture di ogni classe; i risultati ottenuti sono riportati in Tabella 4-1. Questa riporta la percentuale di condotte appartenenti ad una certa classe di materiale, espressa in termini di lunghezza normalizzata rispetto alla lunghezza totale delle condotte e la percentuale di rotture di ogni classe, rispetto alle rotture totali.

Tabella 4-1 Calcolo del tasso di fallanza dei singoli materiali presenti all'interno della tabella *ReteTorino*

Materiale	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
Ghisa grigia	730.30	49.94	2945.00	75.88	4.03
Ghisa sferoidale	397.72	27.20	387.00	9.97	0.97
Acciaio	247.28	16.91	173.00	4.46	0.70
Eternit	59.09	4.04	351.00	9.04	5.94
PEAD	12.61	0.86	9.00	0.23	0.71
PVC	6.53	0.45	4.00	0.10	0.61
Chameroy	4.71	0.32	4.00	0.10	0.85
Cemento armato	2.91	0.20	1.00	0.03	0.34
Ferro	0.83	0.06	2.00	0.05	2.42
Materiali indefiniti	0.04	0.00	0.00	0.00	0.00
Totale	1462.28	100.00	3881.00	100.00	

In Figura 4-2 si riportano, per via grafica, tali risultati e per permettere una valutazione più immediata del tasso di rottura sono stati presi in considerazione soltanto i materiali presenti per più dello 0.3% in lunghezza rispetto al totale della rete.

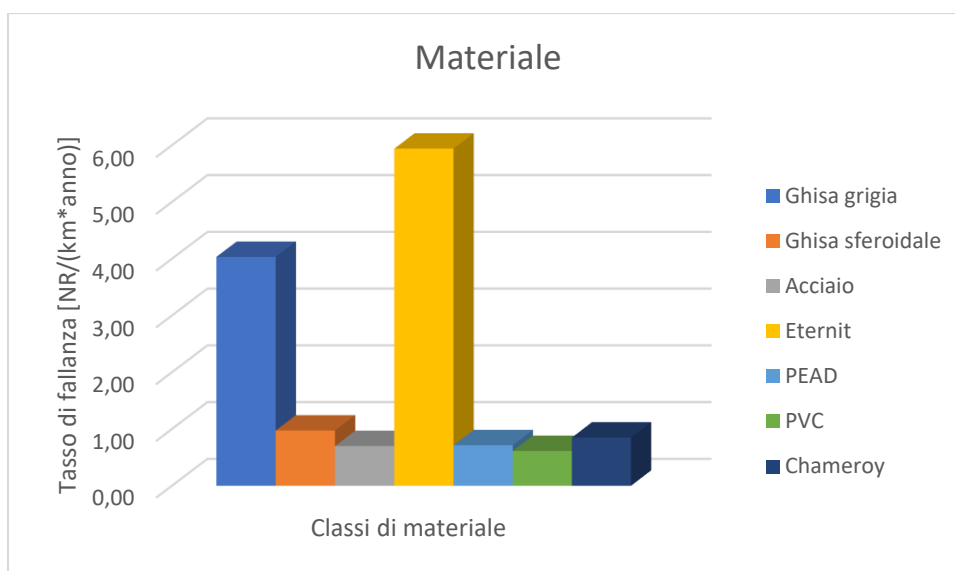


Figura 4-2 Tasso di fallanza dei materiali presenti nella tabella ReteTorino

Ne traspare una maggiore facilità di fallanza delle condotte in eternit e in ghisa grigia. I materiali più presenti in rete sono: la ghisa grigia, la ghisa sferoidale e l'acciaio.

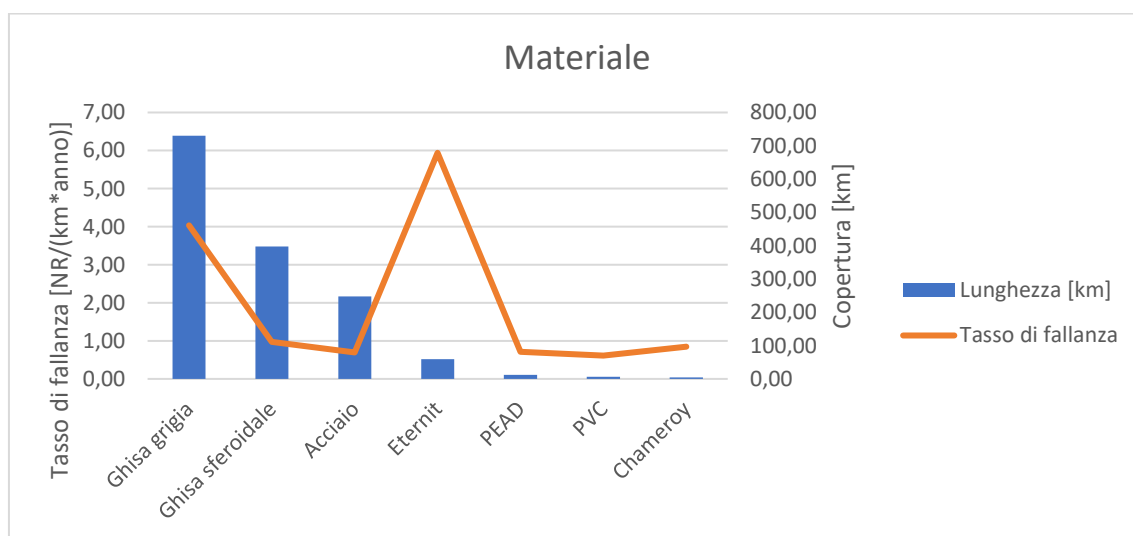


Figura 4-3 Tasso di rottura di rete osservato per i singoli materiali. Copertura offerta dei singoli materiali

Calcolo del tasso di fallanza delle classi di diametro

Allo stesso modo è stato determinato il tasso di fallanza per le diverse classi di diametri. Si riportano i risultati in Tabella 4-2.

Tabella 4-2 Calcolo del tasso di fallanza delle classi di diametro presenti all'interno della tabella ReteTorino

Classe	Diametro [mm]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	$D \leq 100$ mm	589.78	40.32	2777.00	71.55	4.71
2	$100 \text{ mm} < D < 200$ mm	225.44	15.41	693.00	17.86	3.07
3	$D \geq 200$ mm	647.59	44.27	411.00	10.59	0.63
Totale		1462.82	100.00	3881.00	100.00	

La Figura 4-4 descrive l'andamento del tasso di fallanza; i valori ottenuti consentono di affermare che questo aumenta al diminuire del diametro.

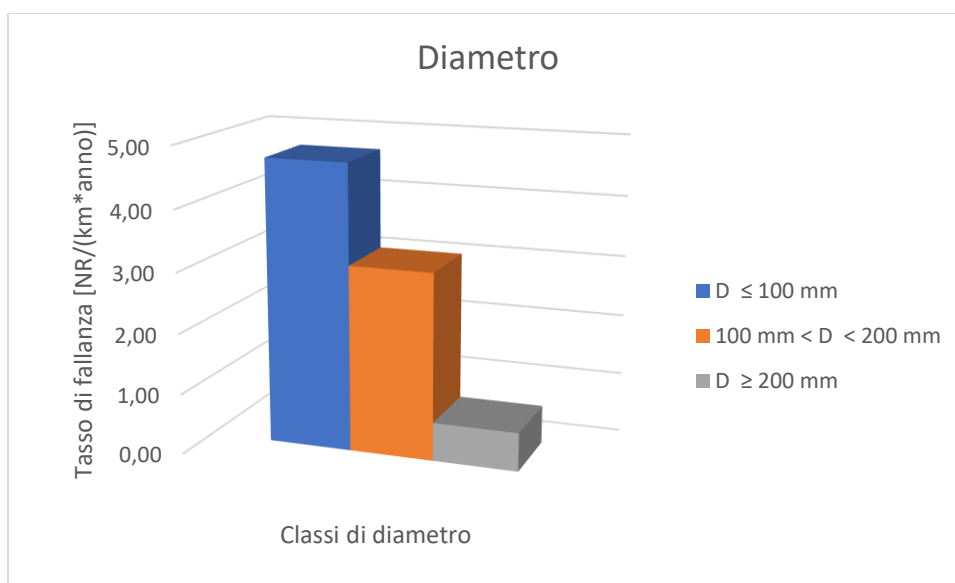


Figura 4-4 Tasso di fallanza delle classi di diametro presenti nella tabella ReteTorino

Come mostrato nel grafico seguente, non si è riscontrata una correlazione proporzionale tra il numero di rotture rilevato per un determinato diametro e la sua estensione.

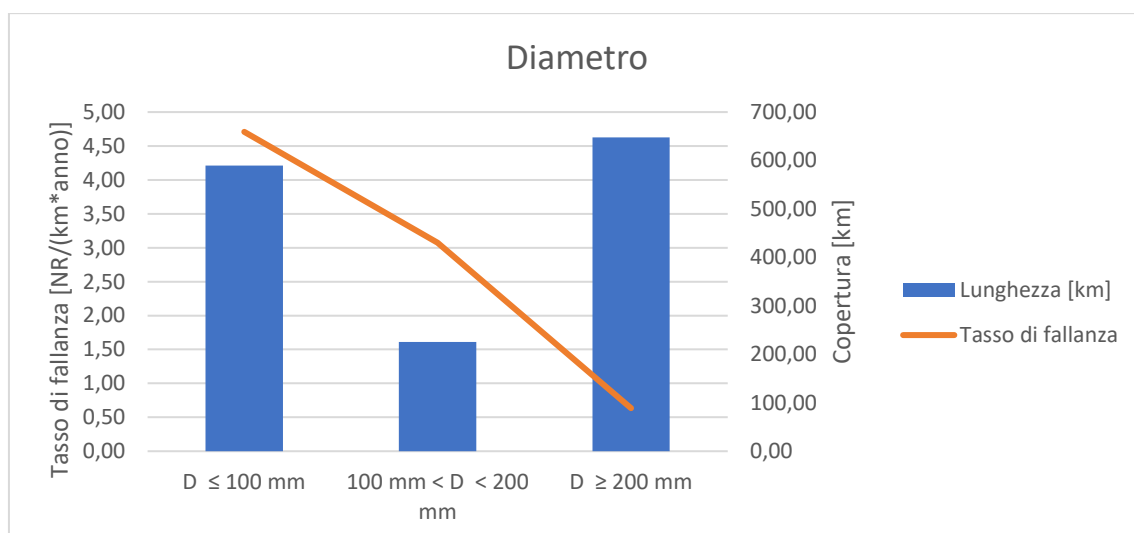


Figura 4-5 Tasso di rottura di rete osservato per le classi di diametro. Copertura offerta delle classi di diametro

Calcolo del tasso di fallanza delle classi relative all'anno di posa

Lo stesso procedimento è stato adottato per le informazioni relative agli anni di posa. Tali informazioni sono state classificate in classi ventennali e in seguito sono state calcolate le lunghezze totali e le occorrenze delle rotture di ogni classe. Il 76% delle condotte presenti in rete sono prive dell'informazione relativa all'anno di posa.

Tabella 4-3 Calcolo del tasso di fallanza delle classi di posa presenti nella tabella ReteTorino

Classe	Anno posa	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	Anno indefinito	1113.12	76.09	3463.00	89.23	3.11
2	1870 - 1890	0.91	0.06	3.00	0.08	3.29
3	1890 - 1910	9.84	0.67	35.00	0.90	3.56
4	1910 - 1930	13.74	0.94	16.00	0.41	1.16
5	1930 - 1950	2.88	0.20	23.00	0.59	7.99
6	1950 - 1970	6.05	0.41	18.00	0.46	2.98
7	1970 - 1990	32.68	2.23	24.00	0.62	0.73
8	1990 - 2010	246.54	16.85	269.00	6.93	1.09
9	2010 →	37.06	2.53	30.00	0.77	0.81
Totale		1462.82	100.00	3881.00	100.00	

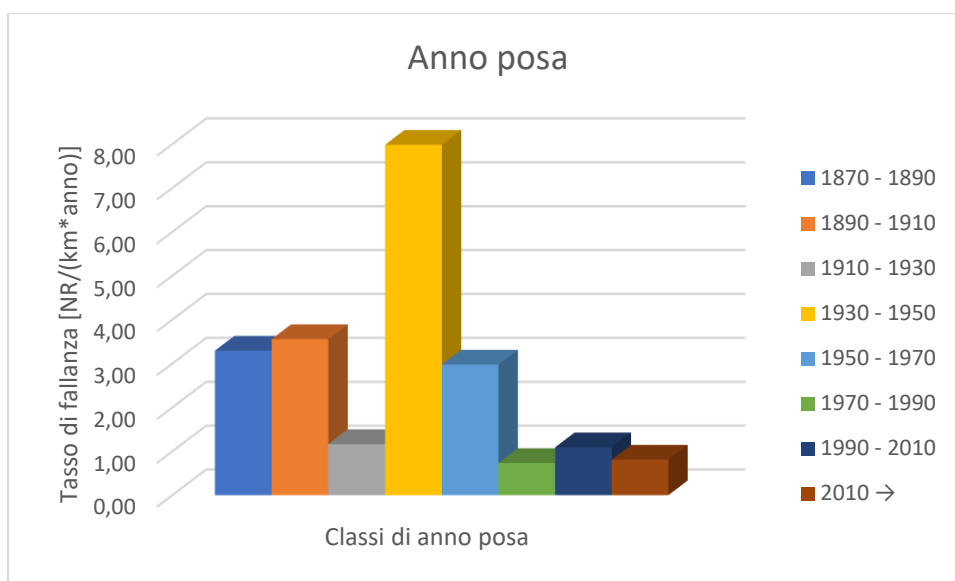


Figura 4-6 Tasso di fallanza delle classi di posa presenti nella tabella ReteTorino

Una percentuale molto elevata di condotte non possiede il dato relativo all'anno di posa, di conseguenza le condotte in questione non saranno presenti nelle tabelle successive, sulle quali verrà applicato il modello di regressione. L'analisi delle classi in cui l'anno di posa è noto, consente di affermare che le condotte più diffuse sono state posate nel ventennio 1990-2010, in cui però non si riscontra il tasso di rottura più alto. Le condotte posate nel ventennio 1930-1950 sono state soggette a molte rotture, malgrado la loro presenza in rete in termini di estensione sia modesta.

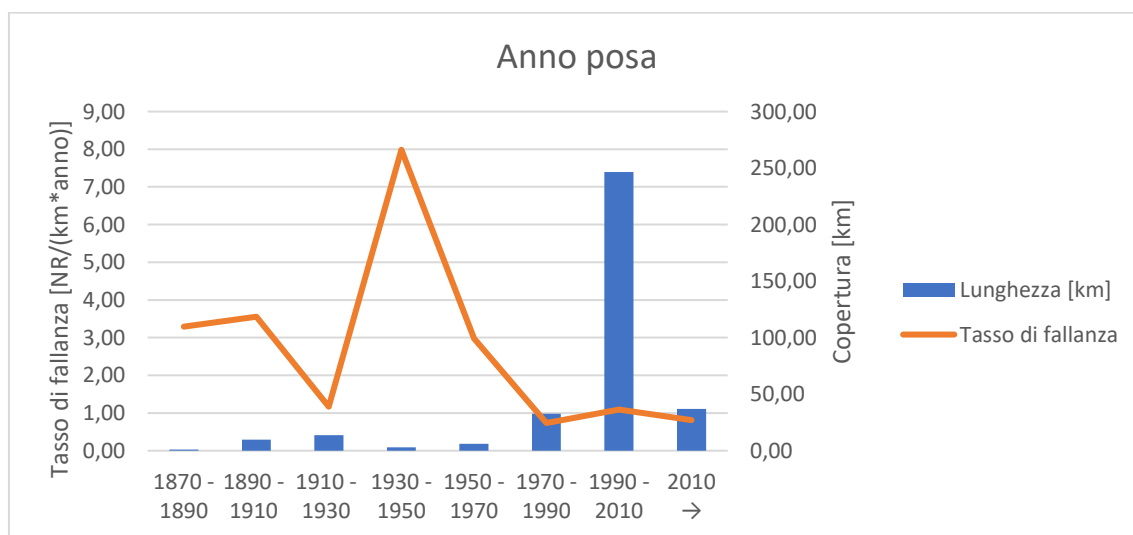


Figura 4-7 Tasso di rottura di rete osservato per le classi di posa. Copertura offerta delle classi di posa

Calcolo del tasso di fallanza delle classi di carico massimo

Sono stati analizzati i carichi massimi. I valori sono stati suddivisi in classi, in cui la variazione degli estremi è di 40 m.

Tabella 4-4 Calcolo del tasso di fallanza delle classi di carico massimo presenti nella tabella ReteTorino

Classe	Carico max [m]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	(0, 40] m	208.01	17.12	579.00	15.18	2.78
2	(40, 80] m	942.00	77.55	3012.00	78.95	3.20
3	(80, 120] m	19.39	1.60	60.00	1.57	3.09
4	(120,160] m	15.18	1.25	67.00	1.76	4.41
5	>160 m	30.15	2.48	97.00	2.54	3.22
Totale		1214.71	100.00	3815.00	100.00	

Il tasso di fallanza mostra piccole variazioni rispetto al valore medio, pari a 3.34 NR/(km*anno), in tutte le classi, eccezion fatta per la quarta. Condotte con carico massimo compreso fra 120 metri e 160 metri presentano il valore più alto di tasso di fallanza, discostandosi di un'unità dal valore medio. La prima e seconda classe sono le più diffuse in termini di copertura in chilometri all'interno della rete.

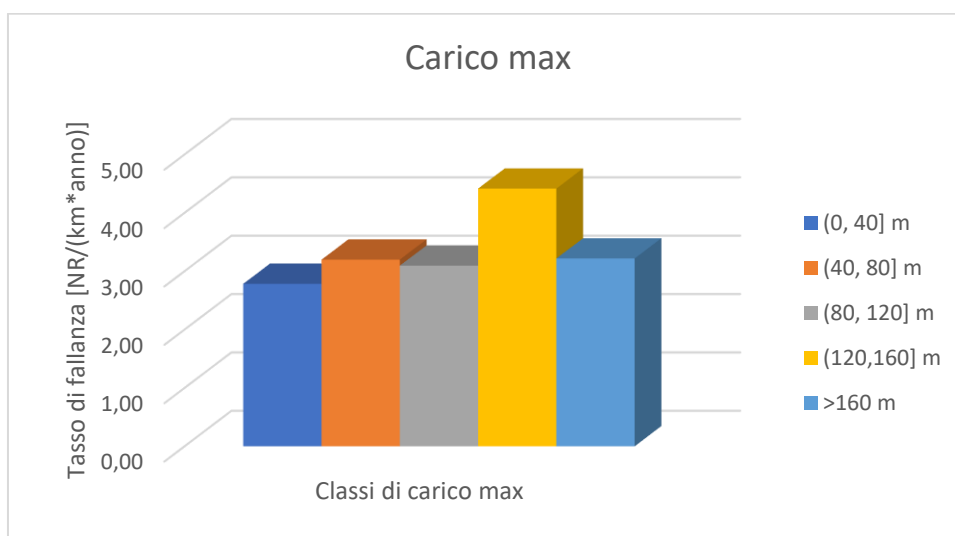


Figura 4-8 Tasso di fallanza delle classi di carico massimo presenti nella tabella ReteTorino

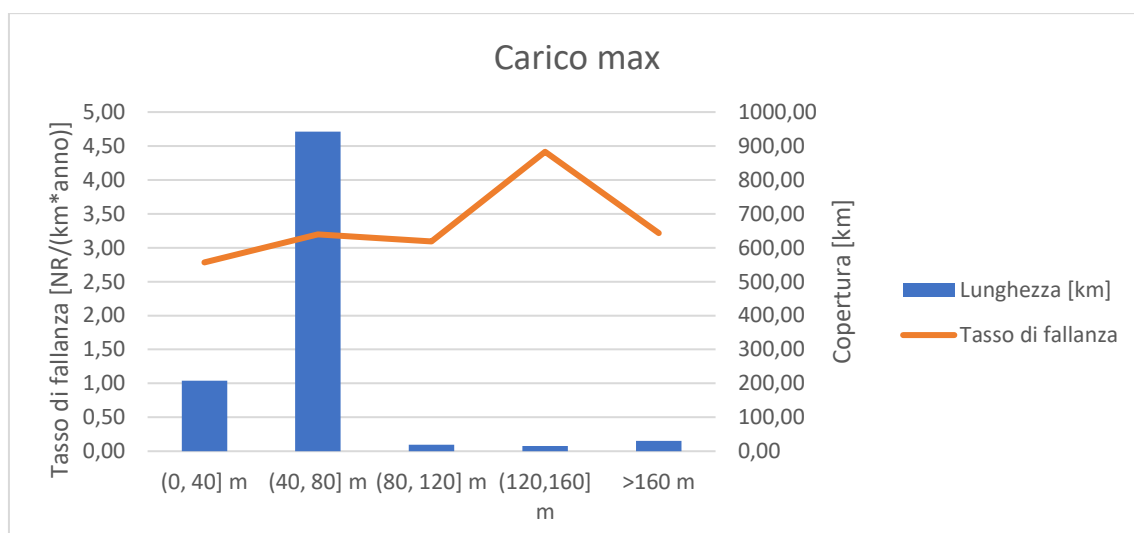


Figura 4-9 Tasso di rottura di rete osservato per le classi di carico massimo. Copertura offerta delle classi di carico massimo

Calcolo del tasso di fallanza delle classi di variazione di carico

Lo stesso procedimento è stato adottato per le variazioni di carico, i cui valori sono stati suddivisi in classi con passo di 10 metri.

Tabella 4-5 Calcolo del tasso di fallanza delle classi di variazione di carico presenti nella tabella ReteTorino

Classe	ΔH [m]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	(0, 10] m	303.27	24.86	723.00	18.92	2.38
2	(10, 20] m	847.04	69.43	2863.00	74.93	3.38
3	(20, 30] m	25.26	2.07	92.00	2.41	3.64
4	(30,40] m	31.06	2.55	79.00	2.07	2.54
5	> 40 m	13.32	1.09	64.00	1.67	4.81
Totale		1219.95	100.00	3821.00	100.00	

La classe più estesa all'interno della rete è la seconda, che rappresenta le condotte con variazione di carico compresa tra 10 metri e 20 metri. Nelle condotte con fluttuazione superiore a 40 metri si riscontra il tasso di rottura più alto, infatti l'improvvisa e cospicua variazione dell'energia pressoria si ripercuote sulle pareti della condotta causando gravi danni.

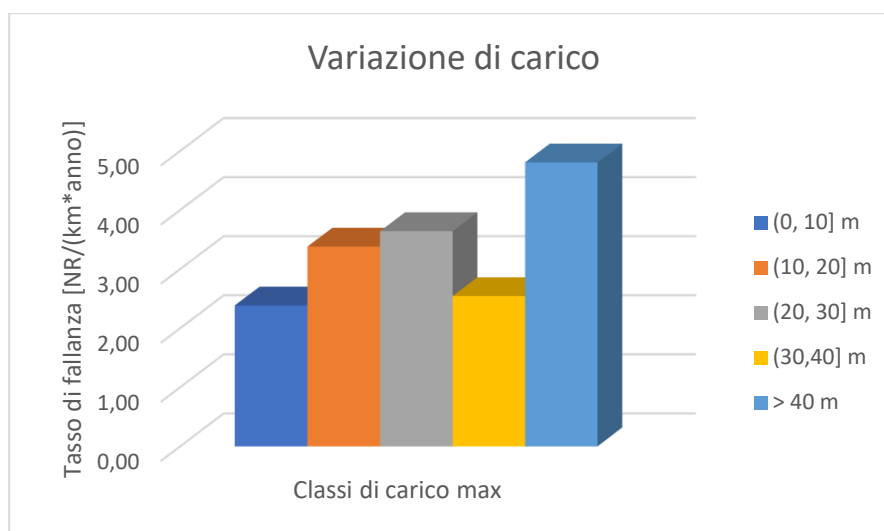


Figura 4-10 Tasso di fallanza delle classi di variazione di carico presenti nella tabella ReteTorino

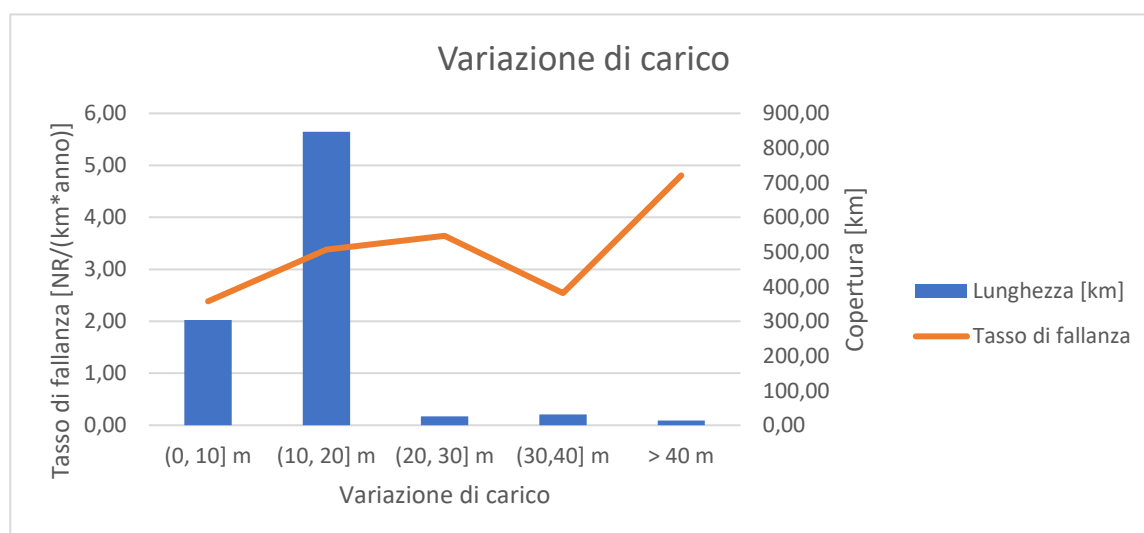


Figura 4-11 Tasso di rottura di rete osservato per le classi di variazioni di carico. Copertura offerta delle classi di carico massimo

4.1.1 Estrazione e analisi di sottotabelle per ogni materiale

Per poter valutare i materiali distintamente, sono state estratte dalla tabella *ReteTorino*, con l'ausilio del software di programmazione *Python*, tante sottotabelle quante sono le tipologie di materiale che presentano una copertura in chilometri maggiore dello 0.5% sul totale della rete. Per ogni materiale è stata valutata la copertura in chilometri, il numero di rotture e le occorrenze delle classi di: diametro, anno di posa, carico massimo e variazione di carico; quindi è stato calcolato il tasso di fallanza. La scelta di analizzare i materiali distintamente è dettata dal diverso comportamento che questi presentano. Si è già anticipato che la rete di Torino è il risultato di uno sviluppo nel tempo, infatti sono presenti condotte costituite da

materiali di più recente impiego ed altri che oggi non sono più in commercio o non vengono più utilizzati (come l'eternit e la ghisa grigia).

Le sottotabelle estratte sono state denominate come segue:

- ❖ *ReteTorino_GhisaGrigia*;
- ❖ *ReteTorino_GhisaSferoidale*;
- ❖ *ReteTorino_Acciaio*;
- ❖ *ReteTorino_Eternit*;
- ❖ *ReteTorino_PEAD*.

Tabella ReteTorino_GhisaGrigia

Si riporta l'analisi delle classi di diametro, anno di posa, carico massimo e variazione di carico.

- **Diametro:**

Tabella 4-6 Calcolo del tasso di fallanza delle classi di diametro presenti nella tabella ReteTorino_GhisaGrigia

Classe	Diametro [mm]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	D ≤ 100	405.98	55.59	2232	75.79	5.50
2	100 < D < 200	153.42	21.01	575	19.52	3.75
3	D ≥ 200	170.90	23.40	138	4.69	0.81
Totale		730.3	100	2945	100	

La ghisa grigia è in assoluto il materiale più diffuso. I diametri inferiori a 100 millimetri caratterizzano per più del 50% la rete e presentano tasso di fallanza più alto. All'aumentare del diametro, il tasso di fallanza diminuisce.

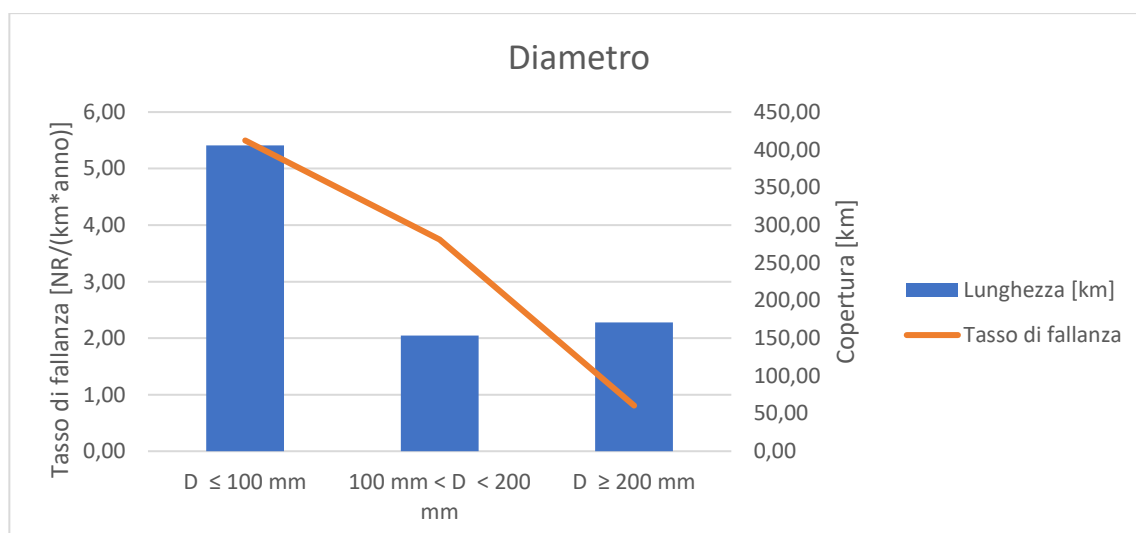


Figura 4-12 Copertura e tasso di fallanza delle classi di diametro della tabella ReteTorino_GhisaGrigia

▪ Anno di posa:

Tabella 4-7 Calcolo del tasso di fallanza delle classi di anno di posa presenti nella tabella ReteTorino_GhisaGrigia

Classe	Anno posa	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	Anno indefinito	700.08	95.86	2855	96.94	4.08
2	1870 - 1890	0.91	0.12	3	0.10	3.29
3	1890 - 1910	9.12	1.25	35	1.19	3.84
4	1910 - 1930	13.10	1.79	11	0.37	0.84
5	1930 - 1950	2.46	0.34	13	0.44	5.28
6	1950 - 1970	1.22	0.17	11	0.37	9.05
7	1970 - 1990	0.62	0.08	4	0.14	6.50
8	1990 - 2010	2.79	0.38	13	0.44	4.66
Totale		730.3	100	2945	100	

Circa il 96% delle condotte in ghisa grigia è priva dell'informazione relativa all'anno di posa e ciò riduce la veridicità dei risultati. La maggior parte delle condotte, di cui è noto il dato, è stata posata prima degli anni '50, periodo a partire dal quale ha inizio la diffusione della ghisa sferoidale.

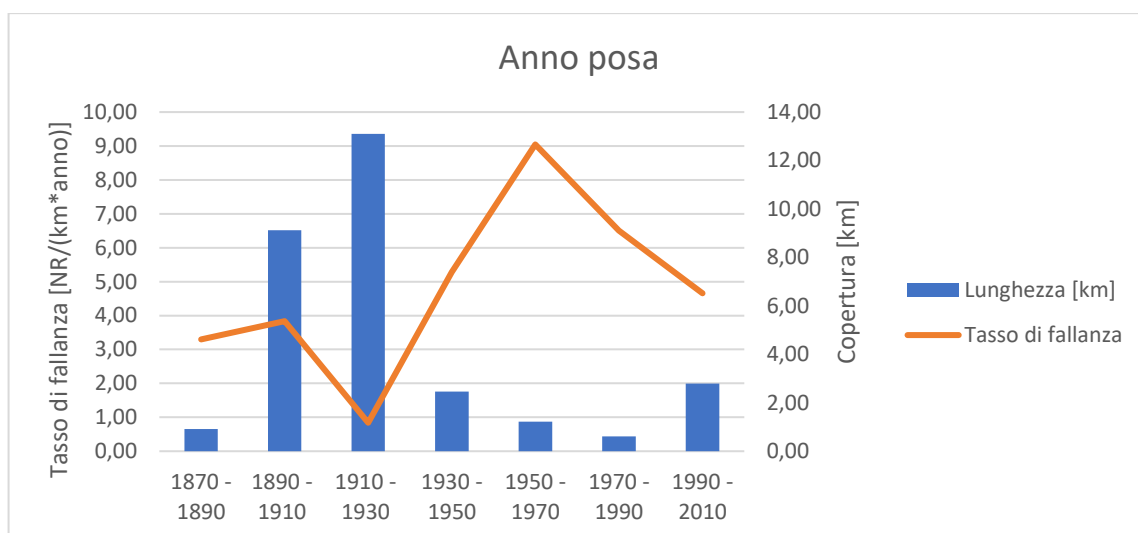


Figura 4-13 Copertura e tasso di fallanza delle classi di anno di posa della tabella ReteTorino_GhisaGrigia

▪ Carico massimo:

Tabella 4-8 Calcolo del tasso di fallanza delle classi di carico massimo presenti nella tabella ReteTorino_GhisaGrigia

Classe	Carico max [m]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	(0, 40] m	99.72	14.58	455.00	15.72	4.56
2	(40, 80] m	563.39	82.40	2313.00	79.92	4.11
3	(80, 120] m	5.57	0.81	40.00	1.38	7.19
4	(120,160] m	5.65	0.83	38.00	1.31	6.72
5	>160 m	9.38	1.37	48.00	1.66	5.12
Totale		683.71	100.00	2894.00	100.00	

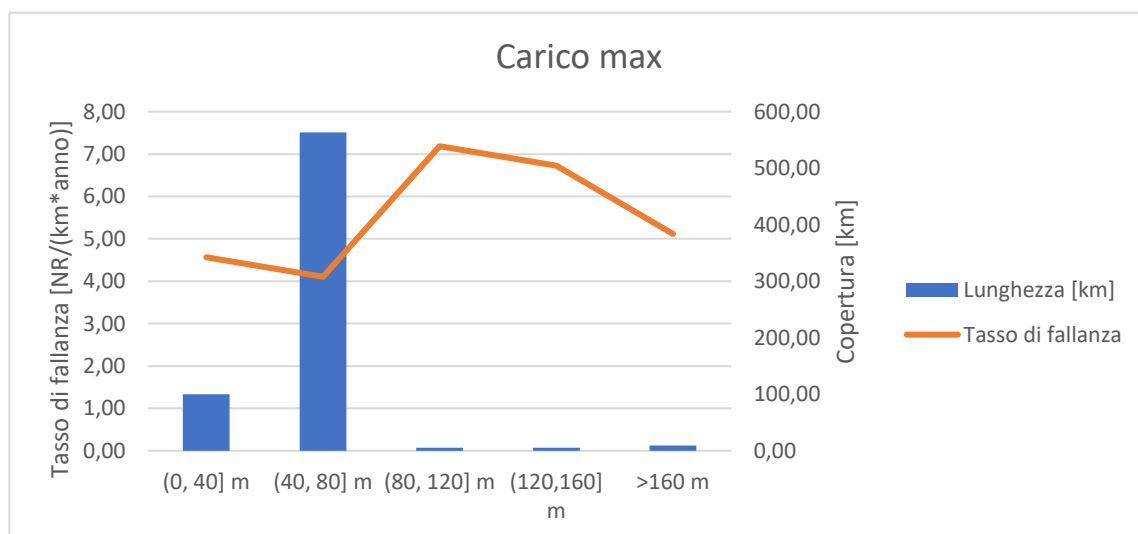


Figura 4-14 Copertura e tasso di fallanza delle classi di carico massimo della tabella ReteTorino_GhisaGrigia

La classe caratterizzata da un tasso di fallanza maggiore è quella tra gli 80 e i 120 metri di carico, presente in rete per meno dell'1%.

- Variazione di carico:

Tabella 4-9 Calcolo del tasso di fallanza delle classi di variazione di carico presenti nella tabella ReteTorino_GhisaGrigia

Classe	ΔH [m]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	(0, 10] m	114.418	16.735	446.000	15.411	3.898
2	(10, 20] m	524.718	76.745	2272.000	78.507	4.330
3	(20, 30] m	15.670	2.292	63.000	2.177	4.020
4	(30,40] m	21.981	3.215	68.000	2.350	3.094
5	> 40 m	6.925	1.013	45.000	1.555	6.498
Totale		683.712	100	2894	100	

Le condotte in ghisa grigia che presentano variazioni di carico maggiori di 40 metri mostrano un elevato numero di guasti e sono quindi più vulnerabili.

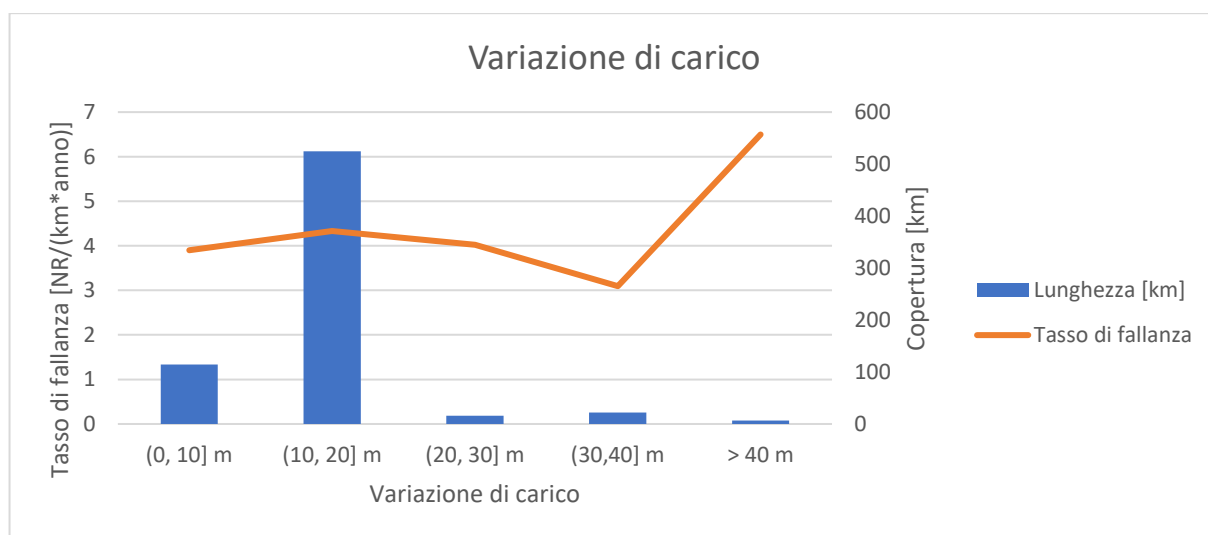


Figura 4-15 Copertura e tasso di fallanza delle classi di variazione di carico della tabella ReteTorino_GhisaGrigia

Tabella ReteTorino GhisaSferoidale

Si riporta l'analisi delle classi di diametro, anno di posa, carico massimo e variazione di carico.

▪ **Diametro:**

Tabella 4-10 Calcolo del tasso di fallanza delle classi di diametro presenti nella tabella ReteTorino_GhisaSferoidale

Classe	Diametro [mm]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	D ≤ 100 mm	131.78	33.13	197	50.90	1.49
2	100 mm < D < 200 mm	142.64	35.86	140	36.18	0.98
3	D ≥ 200 mm	123.31	31.00	50	12.92	0.41
Totale		397.72	100.00	387	100.00	

Anche in questo caso il tasso di fallanza diminuisce all'aumentare del diametro. La percentuale maggiore di tratti presenti in rete fanno riferimento a diametri compresi tra i 100 e i 200 millimetri.

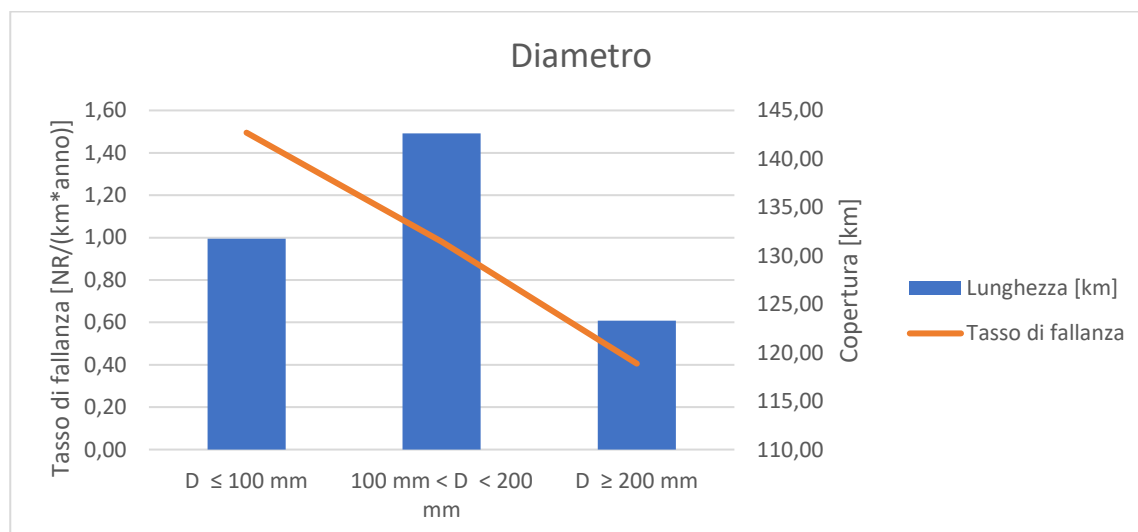


Figura 4-16 Copertura e tasso di fallanza delle classi di diametro della tabella ReteTorino_GhisaSferoidale

▪ **Anno di posa:**

Tabella 4-11 Calcolo del tasso di fallanza delle classi di anno di posa presenti nella tabella ReteTorino_GhisaSferoidale

Classe	Anno posa	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	Anno indefinito	142.44	35.81	132	34.11	0.93
2	1950 - 1970	0.17	0.04	0	0.00	0.00
3	1970 - 1990	19.05	4.79	13	3.36	0.68
4	1990 - 2010	202.35	50.88	212	54.78	1.05
5	2010 →	33.71	8.48	30	7.75	0.89
Totale		397.721	100	387	100	

La ghisa sferoidale può essere definito un materiale di moderno impiego, circa il 50% delle condotte sono state posate a partire dal 1990 e non si hanno informazioni di condotte posate

prima del 1950. Negli ultimi anni è aumentata l'attenzione nel riportare le caratteristiche delle condotte, per questo motivo il 65% dei tratti dispone di questa informazione. La Figura 4-17 mostra come le condotte più recenti, che sono le più diffuse all'interno della rete, presentino il tasso di fallanza più alto.

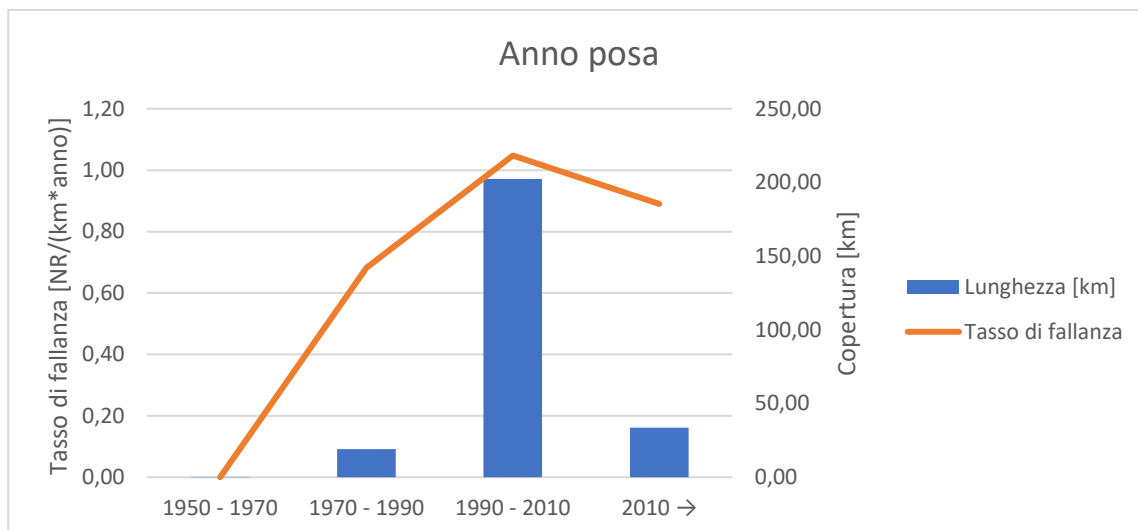


Figura 4-17 Copertura e tasso di fallanza delle classi di anno di posa della tabella ReteTorino_GhisaSferoidale

▪ Carico massimo:

Tabella 4-12 Calcolo del tasso di fallanza delle classi di carico massimo presenti nella tabella ReteTorino_GhisaSferoidale

Classe	Carico max [m]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	(0, 40] m	73.17	22.92	69.00	18.21	0.94
2	(40, 80] m	219.37	68.71	256.00	67.55	1.17
3	(80, 120] m	7.11	2.23	11.00	2.90	1.55
4	(120,160] m	6.79	2.13	11.00	2.90	1.62
5	>160 m	12.83	4.02	32.00	8.44	2.49
Totale		319.26	100.00	379.00	100.00	

Il 68% delle condotte in ghisa sferoidale presenta carichi compresi tra 40 e 80 metri. Il tasso di fallanza massimo è registrato per condotte aventi carichi maggiori di 160 metri.

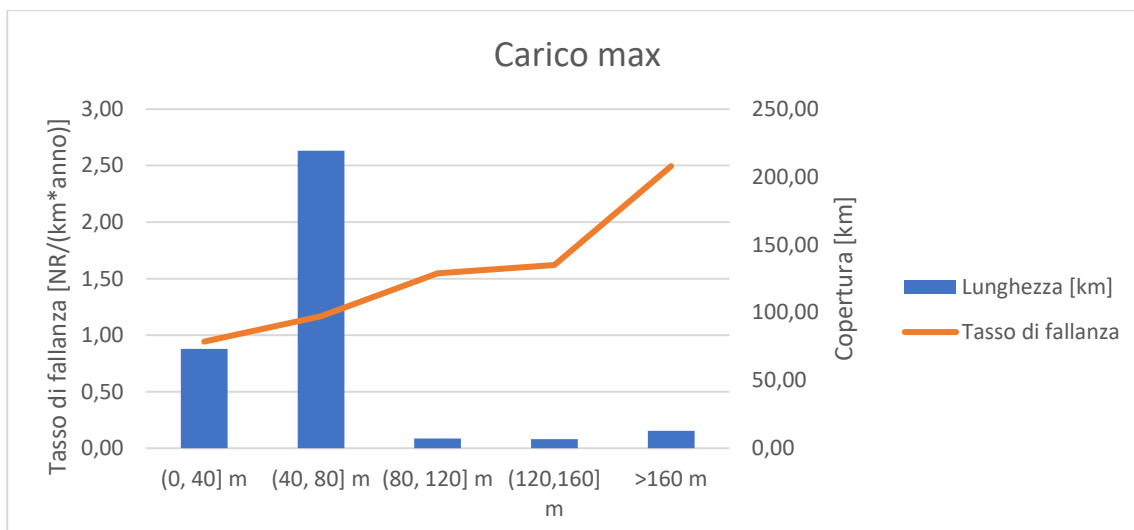


Figura 4-18 Copertura e tasso di fallanza delle classi di carico massimo della tabella ReteTorino_GhisaSferoidale

▪ Variazione di carico:

Tabella 4-13 Calcolo del tasso di fallanza delle classi di variazione di carico della tabella ReteTorino_GhisaSferoidale

Classe	ΔH [m]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	(0, 10] m	102.679	32.010	92.000	24.211	0.896
2	(10, 20] m	203.133	63.327	259.000	68.158	1.275
3	(20, 30] m	6.748	2.104	21.000	5.526	3.112
4	(30,40] m	5.387	1.679	1.000	0.263	0.186
5	> 40 m	2.822	0.880	7.000	1.842	2.481
Totale		320.769	100	380	100	

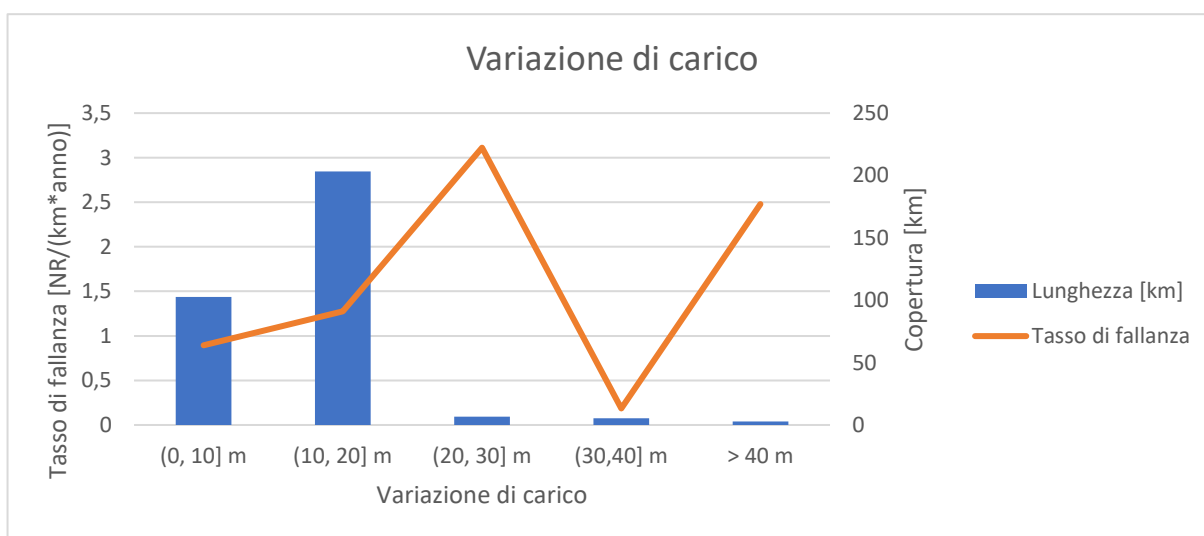


Figura 4-19 Copertura e tasso di fallanza delle classi di variazione di carico della tabella ReteTorino_GhisaSferoidale

I valori di tasso di fallanza ottenuti per ogni classe si discostano di circa un'unità l'uno dall'altro. I valori di picco sono associati alle classi meno estese in termini di chilometri.

Tabella ReteTorino Acciaio

L'acciaio è il terzo materiale più presente in rete.

Si riporta l'analisi delle classi di diametro, anno di posa, carico massimo e variazione di carico delle condotte costituite da acciaio.

- Diametro:

Tabella 4-14 Calcolo del tasso di fallanza delle classi di diametro presenti nella tabella ReteTorino_Acciaio

Classe	Diametro [mm]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	D ≤ 100 mm	19.30	7.80	61	35.26	3.16
2	100 mm < D < 200 mm	20.84	8.43	29	16.76	1.39
3	D ≥ 200 mm	207.15	83.77	83	47.98	0.40
Totale		247.28	100.00	173	100.00	

L'84% delle condotte in acciaio ha un diametro maggiore di 200 millimetri. Anche per questa tipologia di materiale, il tasso di fallanza diminuisce all'aumentare del diametro.

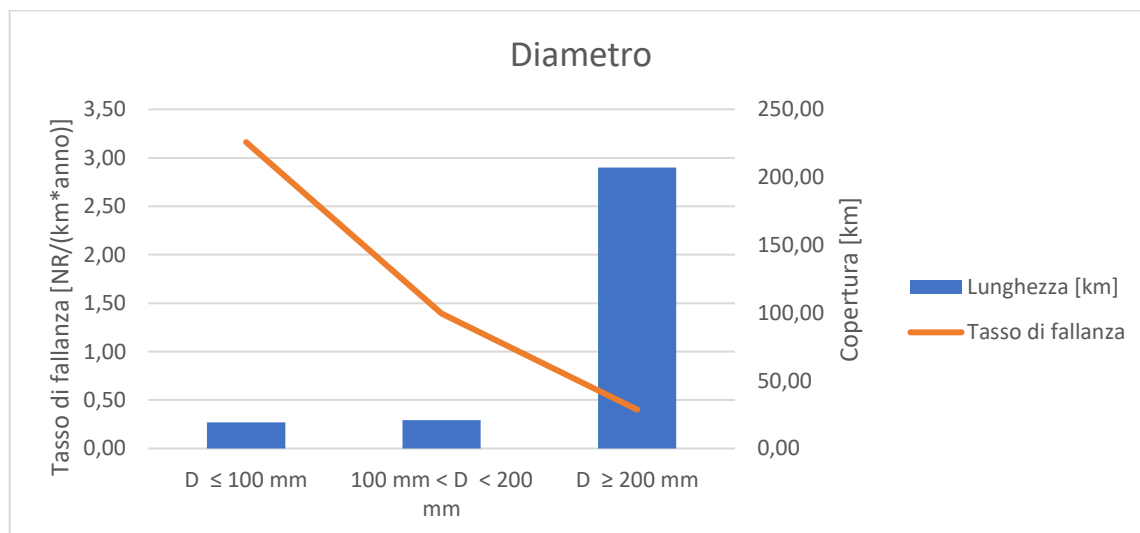


Figura 4-20 Copertura e tasso di fallanza delle classi di diametro della tabella ReteTorino_Acciaio

▪ Anno di posa:

Tabella 4-15 Calcolo del tasso di fallanza delle classi di anno di posa presenti nella tabella ReteTorino_Acciaio

Classe	Anno posa	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1.00	Anno indefinito	188.58	76.26	125	72.25	0.66
2.00	1870 - 1890	0.00	0.00	0	0.00	0.00
3.00	1890 - 1910	0.01	0.00	0	0.00	0.00
4.00	1910 - 1930	0.28	0.11	0	0.00	0.00
5.00	1930 - 1950	0.00	0.00	0	0.00	0.00
6.00	1950 - 1970	4.66	1.88	7	4.05	1.50
7.00	1970 - 1990	11.42	4.62	3	1.73	0.26
8.00	1990 - 2010	39.06	15.79	38	21.97	0.97
9.00	2010 →	3.28	1.33	0	0.00	0.00
Totale		247.28	100.00	173.00	100.00	

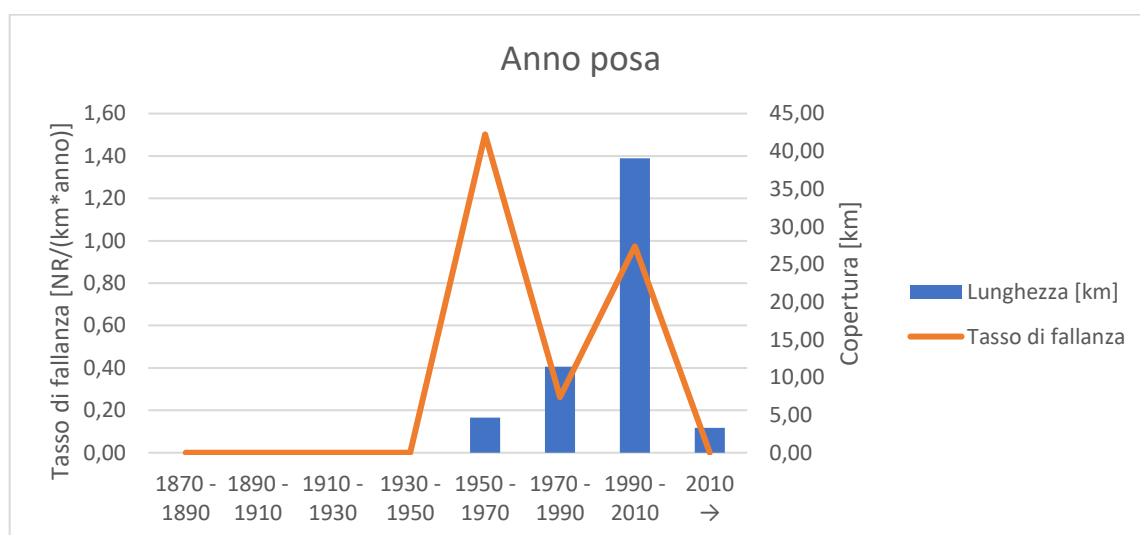


Figura 4-21 Copertura e tasso di fallanza delle classi di anno di posa della tabella ReteTorino_Acciaio

L'utilizzo dell'acciaio risale al ventennio 1950-1970, che è caratterizzato dal valore del tasso di fallanza più elevato. Le condotte sono state posate per la maggiore negli anni 1990-2010, che presentano il secondo valore più alto di tasso di fallanza. Nonostante il suo utilizzo sia recente, circa il 76% delle condotte non presenta questo dato.

- Carico massimo:

Tabella 4-16 Calcolo del tasso di fallanza delle classi di carico massimo presenti nella tabella ReteTorino_Acciaio

Classe	Carico max [m]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	(0, 40] m	22.061	14.195	8	4.678	0.363
2	(40, 80] m	117.047	75.311	127	74.269	1.085
3	(80, 120] m	6.195	3.986	5	2.924	0.807
4	(120,160] m	2.676	1.722	18	10.526	6.726
5	>160 m	7.439	4.786	13	7.602	1.748
Totale		155.418	100	171	100	

Il 75% delle condotte in acciaio presenta carico massimo compreso tra i 40 e gli 80 metri, mentre il tasso di fallanza più elevato caratterizza le condotte con valori di carico massimo compresi tra 120 e 160 metri. Il valore del tasso di fallanza sembra aumentare all'aumentare del carico con picco massimo in corrispondenza della quarta classe.

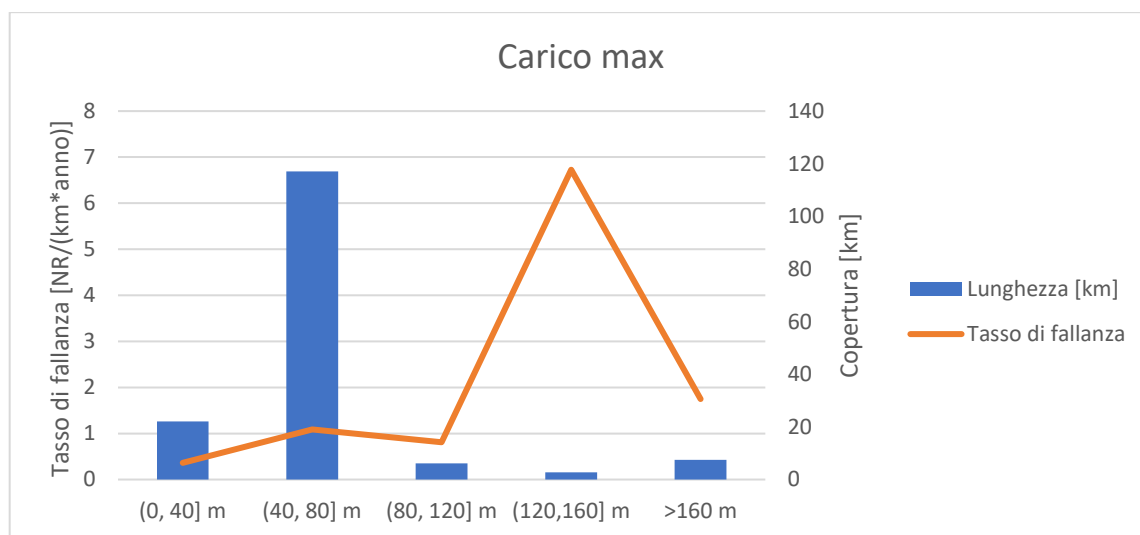


Figura 4-22 Copertura e tasso di fallanza delle classi di carico massimo della tabella ReteTorino_Acciaio

- Variazione di carico:

Tabella 4-17 Calcolo del tasso di fallanza delle classi di variazioni di carico presenti nella tabella ReteTorino_Acciaio

Classe	ΔH [m]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	(0, 10] m	64.201	41.204	59.000	34.503	0.919
2	(10, 20] m	84.816	54.434	97.000	56.725	1.144
3	(20, 30] m	1.68	1.078	0.000	0.000	0.000
4	(30,40] m	2.654	1.703	7.000	4.094	2.638
5	> 40 m	2.462	1.580	8.000	4.678	3.249
Totale		155.813	100	171	100	

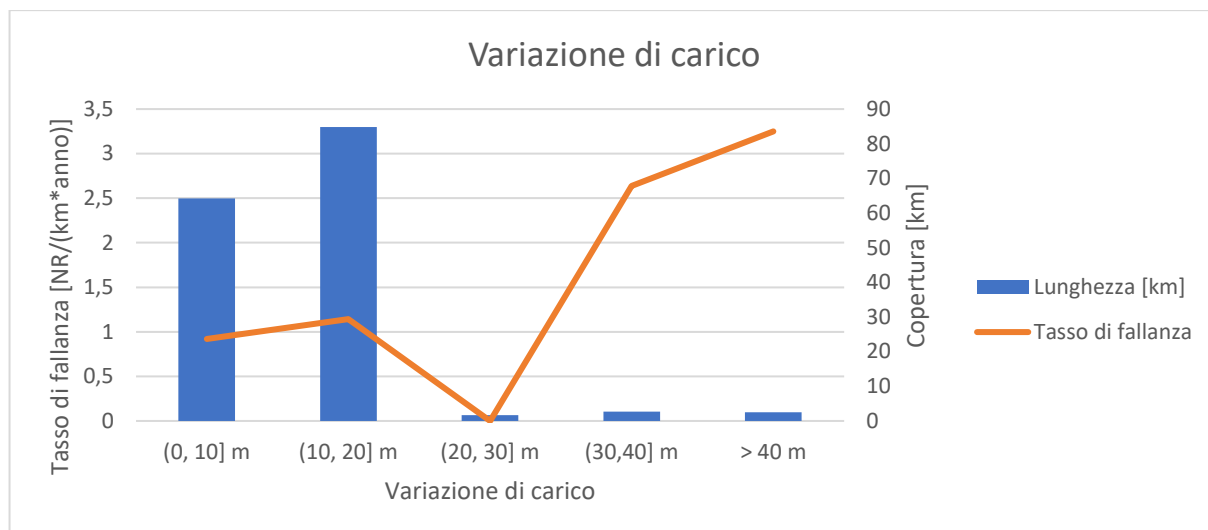


Figura 4-23 Copertura e tasso di fallanza delle classi di variazioni di carico della tabella ReteTorino_Acciaio

Circa la totalità delle condotte in acciaio, di cui sono note le caratteristiche, presentano variazioni di carico comprese fra gli estremi della prima e seconda classe; in queste sono stati registrati il numero maggiore di guasti.

Tabella ReteTorino Eternit

L'eternit, seppur la Normativa abbia imposto la cessazione del suo impiego sin dal 1992, è il quarto materiale più presente in rete.

Si riporta l'analisi delle classi di diametro, anno di posa, carico massimo e variazione di carico delle condotte.

▪ Diametro:

Tabella 4-18 Calcolo del tasso di fallanza delle classi di diametro presenti nella tabella ReteTorino_Eternit

Classe	Diametro [mm]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	D ≤ 100 mm	22.73	38.46	268	76.35	11.79
2	100 mm < D < 200 mm	8.13	13.75	74	21.08	9.11
3	D ≥ 200 mm	28.24	47.79	9	2.56	0.32
Totale		59.09	100.00	351.00	100.00	

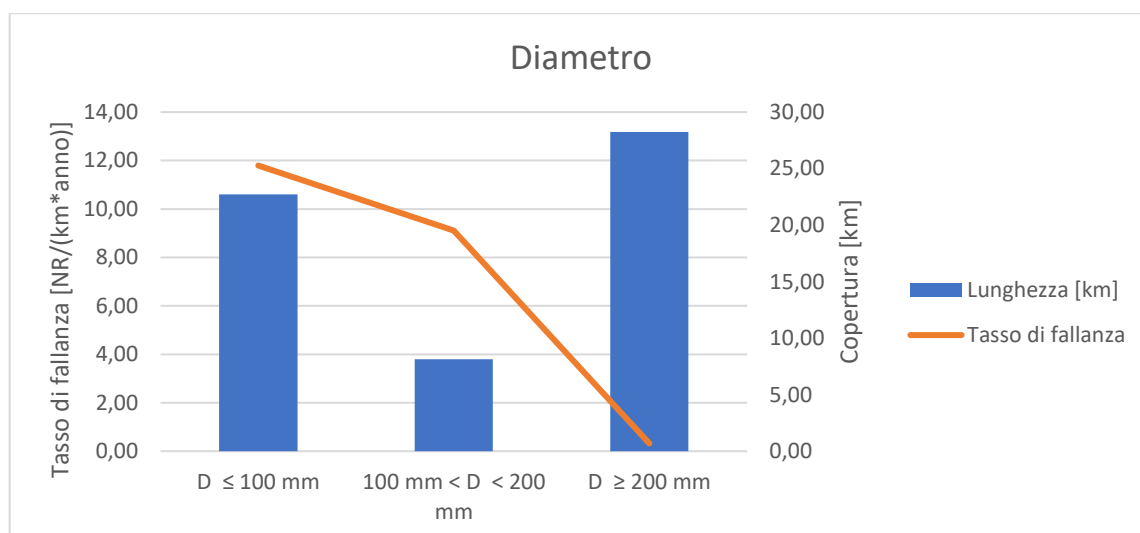


Figura 4-24 Copertura e tasso di fallanza delle classi di diametro della tabella ReteTorino_Eternit

Il tasso di rottura diminuisce all'aumentare del diametro. La maggioranza dei tratti in cemento amianto presenta diametri superiori a 200 millimetri.

- Anno di posa:

Si nota come le percentuali di utilizzo tendano a diminuire in seguito al divieto di commercializzazione e produzione dei prodotti in eternit.

Tabella 4-19 Calcolo del tasso di fallanza delle classi di anno di posa presenti nella tabella ReteTorino_Eternit

Classe	Anno posa	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1.00	Anno indefinito	57.34	97.04	336.00	95.73	5.86
2.00	1910-1930	0.31	0.52	5.00	1.42	16.13
3.00	1930 - 1950	0.63	1.06	10.00	2.85	15.92
4.00	1970 - 1990	0.43	0.73	0.00	0.00	0.00
5.00	1990-2010	0.37	0.63	0.00	0.00	0.00
6.00	2010 →	0.01	0.02	0.00	0.00	0.00
Totale		59.09	100.00	351.00	100.00	

Le condotte più recenti presentano un tasso di fallanza nullo, mentre quelle più vecchie sono caratterizzate da tassi molto elevati. I risultati non sono attendibili poiché la percentuale in chilometri di condotte che dispongono di quest'informazione è pari solo al 3%.

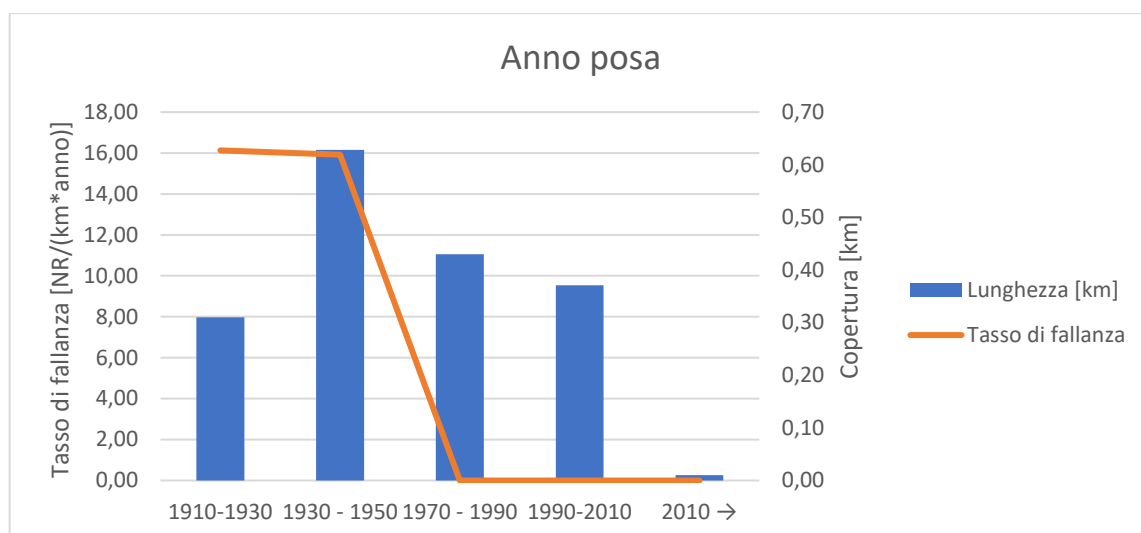


Figura 4-25 Copertura e tasso di fallanza delle classi di anno di posa della tabella ReteTorino_Eternit

▪ Carico massimo:

I risultati mostrano un picco del tasso di fallanza per carichi massimi compresi tra 40 e 80 metri, valori che caratterizzano per circa il 70% la copertura della rete.

Tabella 4-20 Calcolo del tasso di fallanza delle classi di carico massimo presenti nella tabella ReteTorino_Eternit

Classe	Carico max [m]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	(0, 40] m	17.39	41.62	79.00	23.51	4.54
2	(40, 80] m	23.90	57.19	257.00	76.49	10.76
3	>160 m	0.50	1.19	0.00	0.00	0.00
Totale		41.78	100.00	336	100.00	

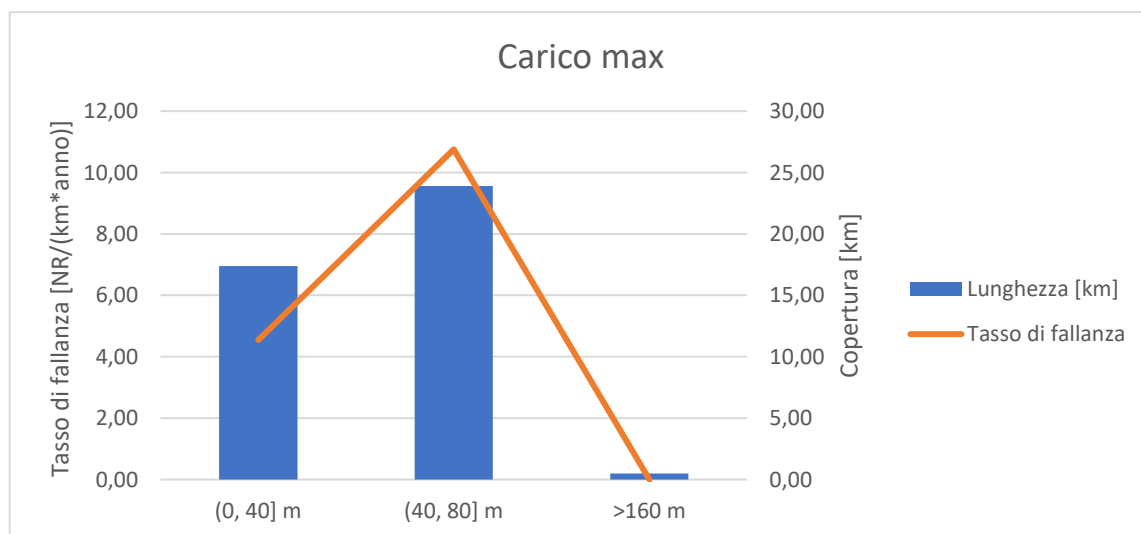


Figura 4-26 Copertura e tasso di fallanza delle classi di carico massimo della tabella ReteTorino_Eternit

▪ Variazione di carico

Tabella 4-21 Calcolo del tasso di fallanza delle classi di variazione di carico presenti nella tabella ReteTorino_Eternit

Classe	ΔH [m]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	(0, 10] m	13.181	30.377	118.000	33.618	8.952
2	(10, 20] m	27.546	63.483	222.000	63.248	8.059
3	(20, 30] m	1.124	2.590	8.000	2.279	7.117
4	(30,40] m	0.833	1.920	3.000	0.855	3.601
5	> 40 m	0.707	1.629	0.000	0.000	0.000
Totale		43.391	100	351	100	

Sono stati calcolati tassi di rottura elevati per le classi più estese.

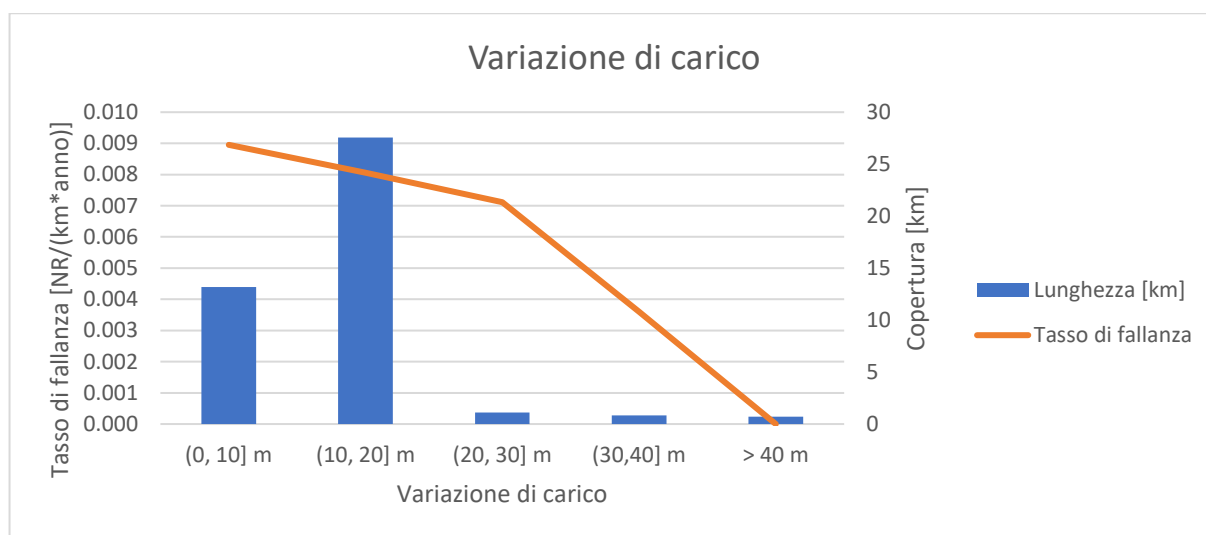


Figura 4-27 Copertura e tasso di fallanza delle classi di variazioni di carico della tabella ReteTorino_Eternit

Tabella ReteTorino PEAD

Come nei casi precedenti, questa tabella è stata ottenuta estraendo tutte le condotte in PEAD dalla tabella *ReteTorino*. Si analizzano i campi diametro, anno di posa, carico massimo e variazione di carico.

Il PEAD è un materiale utilizzato per diramazioni di breve entità. Durante l'elaborazione delle tabelle, molte informazioni su questo materiale sono state eliminate, perché lunghezze contenute, inferiori a 10 metri, non sono sufficientemente significative ai fini della stima del modello.

- Diametro:

Tabella 4-22 Calcolo del tasso di fallanza delle classi di diametro presenti nella tabella ReteTorino_PEAD

Classe	Diametro [mm]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	D ≤ 100 mm	6.105	48.05	8	88.89	1.31
2	100 mm < D < 200 mm	0.486	3.82	1	11.11	2.06
3	D ≥ 200 mm	6.114	48.12	0	0.00	0.00
Totale		12.705	100	9	100.00	

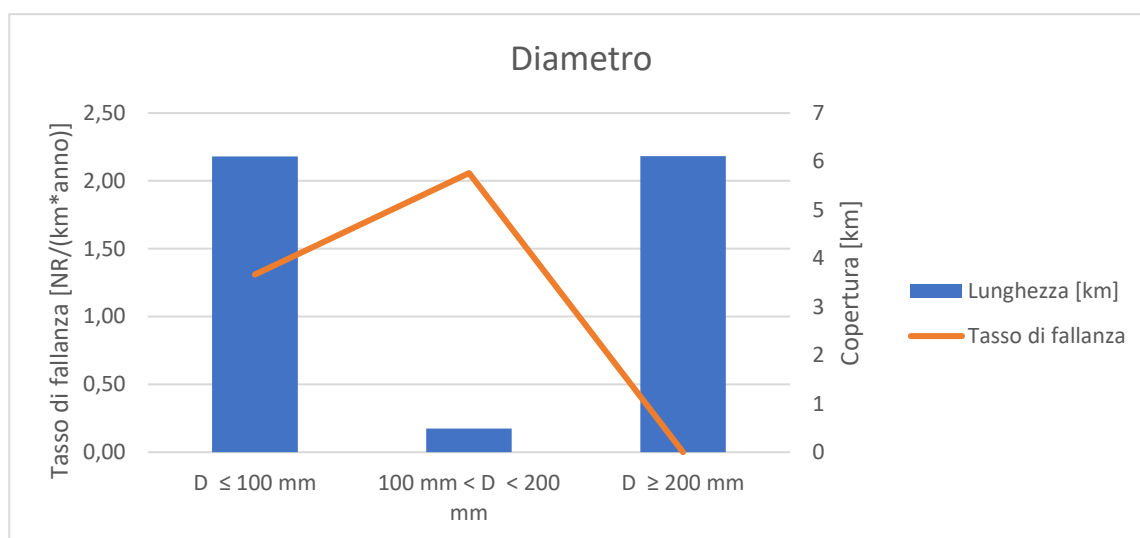


Figura 4-28 Copertura e tasso di fallanza delle classi di diametro della tabella ReteTorino_PEAD

I risultati sono stati ottenuti non tenendo conto di tutti gli elementi presenti in rete. Il tasso di rottura ha un trend non costante all'aumentare del diametro, infatti a differenza degli altri materiali analizzati, non diminuisce all'aumentare del diametro.

- Anno di posa:

Tabella 4-23 Calcolo del tasso di fallanza delle classi di anno di posa presenti nella tabella ReteTorino_PEAD

Classe	Anno posa	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	Anno indefinito	9.47	74.57	1	11.11	0.11
2	1970 - 1990	1.17	9.18	4	44.44	3.43
3	1990 - 2000	0.72	5.67	0	0.00	0.00
4	2000 - 2010	1.28	10.09	4	44.44	3.12
5	2010 →	0.06	0.49	0	0	0.00
Totale		12.71	100.00	9	100	

Il polietilene è un materiale di recente utilizzo, infatti non risultano tratti antecedenti al 1970. Tuttavia, solo per il 25% delle condotte è stata riportata l'informazione sull'anno di posa dagli operatori.

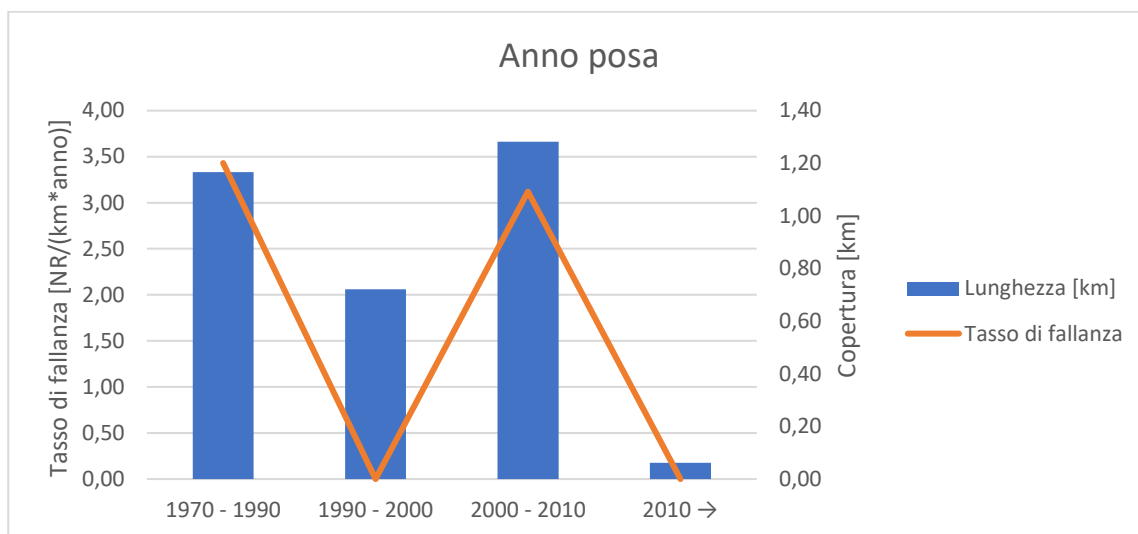


Figura 4-29 Copertura e tasso di fallanza delle classi di anno di posa della tabella ReteTorino_PEAD

▪ Carico massimo:

Tabella 4-24 Calcolo del tasso di fallanza delle classi di carico massimo presenti nella tabella ReteTorino_PEAD

Classe	Carico max [m]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	(0, 40] m	0.37	5.56	1.00	11.11	2.71
2	(40, 80] m	5.47	82.50	2.00	22.22	0.37
3	(80, 120] m	0.48	7.19	4.00	44.44	8.39
4	>160 m	0.32	4.75	2.00	22.22	6.35
Totale		6.64	100.00	9.00	100.00	

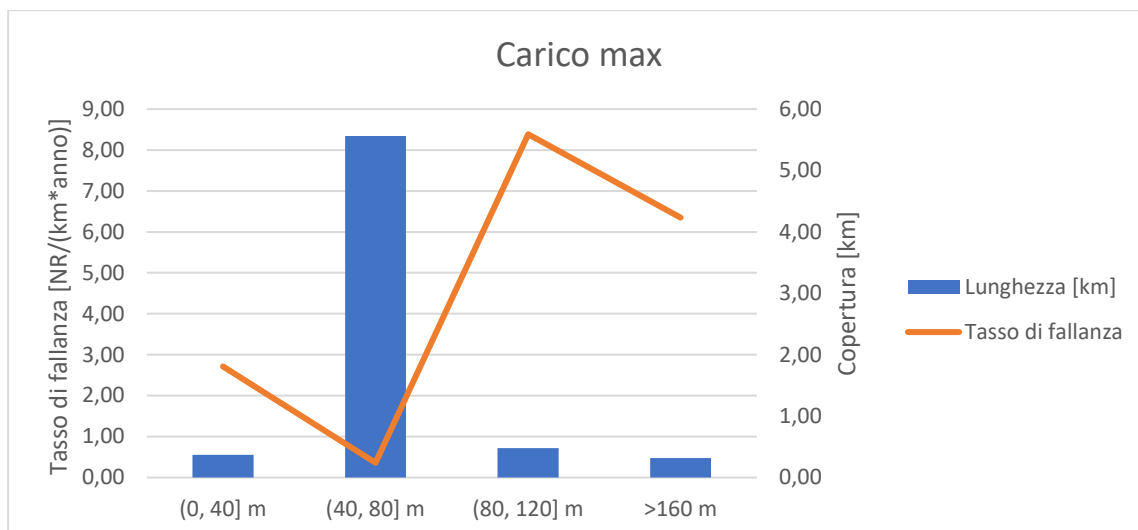


Figura 4-30 Copertura e tasso di fallanza delle classi di carico massimo della tabella ReteTorino_PEAD

Il tasso di rottura cresce notevolmente per valori di carico massimo superiori a 80 metri. L'87% delle condotte è caratterizzato da carichi massimi compresi tra 40 e 80 metri.

- Variazione di carico:

Tabella 4-25 Calcolo del tasso di fallanza delle classi di variazione di carico presenti nella tabella ReteTorino_PEAD

Classe	ΔH [m]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	(0, 10] m	2.067	31.153	4	44.444	1.935
2	(10, 20] m	4.125	62.170	3	33.333	0.727
3	(20, 30] m	0.04	0.603	0	0.000	0.000
4	(30,40] m	0.098	1.477	0	0.000	0.000
5	> 40 m	0.305	4.597	2	22.222	6.557
Totale		6.635	100	9	100	

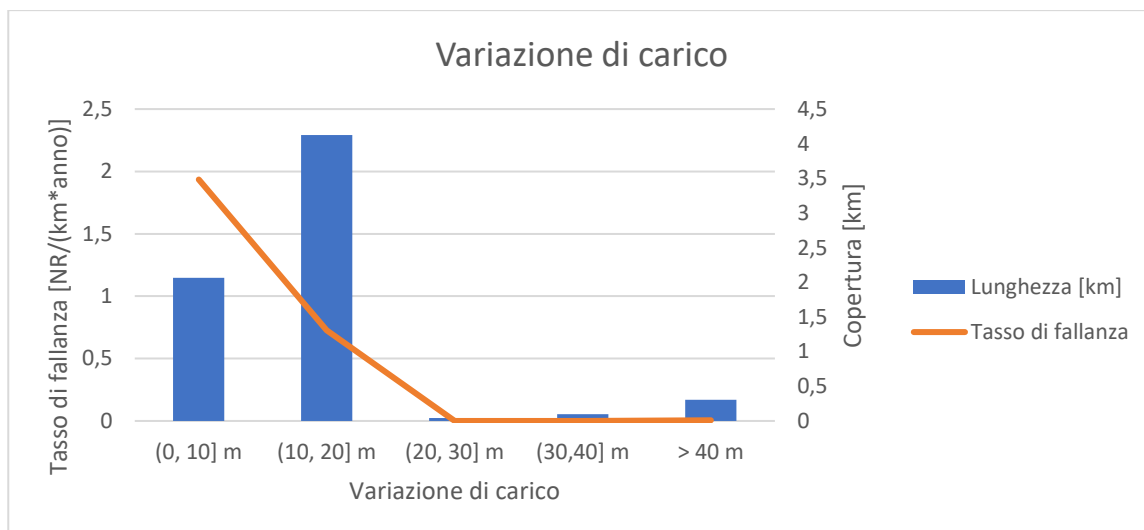


Figura 4-31 Copertura e tasso di fallanza delle classi di variazione di carico della tabella ReteTorino_PEAD

Per i dati a disposizione le condotte più vulnerabili, che presentano tasso di rottura maggiore, sono caratterizzate da variazione di carico maggiore di 40 metri.

4.2 ANALISI DELLA TABELLA REGRESSIONECONDOTTE

La tabella *RegressioneCondotte* è la tabella di partenza, contenente i fattori strutturali e operativi, materiale, diametro, anno di posa, carico massimo e variazione di carico, sulla quale è possibile applicare i modelli di regressione. Poiché questa è costituita di tutte e sole le condotte della tabella *ReteTorino* che presentano un quadro completo di informazioni nei campi precedentemente elencati, presenterà una percentuale di rottura notevolmente più bassa.

Calcolo del tasso di fallanza delle classi di materiale

Come nell'analisi della tabella *ReteTorino*, per ogni materiale è stata calcolata la lunghezza complessiva delle condotte, il numero di guasti e il tasso di fallanza. In accordo con quanto fatto precedentemente per permettere una valutazione più immediata del tasso di rottura sono stati presi in considerazione soltanto i materiali presenti per più dello 0.5% in lunghezza rispetto all'estensione della rete.

Tabella 4-26 Calcolo del tasso di fallanza dei singoli materiali presenti all'interno della tabella *RegressioneCondotte*

Materiali	N° materiale	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
Ghisa grigia	2	24.17	8.46	59	55.66	2.44
Ghisa sferoidale	4	210.93	73.83	32	30.19	0.15
Eternit	1	1.46	0.51	10	9.43	6.85
Acciaio	7	45.14	15.80	5	4.72	0.11
PEAD	6	3.19	1.12	0	0.00	0.00
Totale		285.68	100	106	100	

L'eliminazione di tutte le condotte prive di informazioni fa sì che il materiale predominante sia la ghisa sferoidale, in contrasto con i risultati della tabella *ReteTorino* in cui occupa il secondo posto, dopo la ghisa grigia. Tale risultato è dovuto al più recente impiego della ghisa sferoidale e alla maggiore frequenza negli ultimi anni di riportare le informazioni. Di conseguenza, materiali più recenti sono più frequenti nella tabella *RegressioneCondotte* per il semplice fatto che le condotte sono caratterizzate da tutte le informazioni utili.

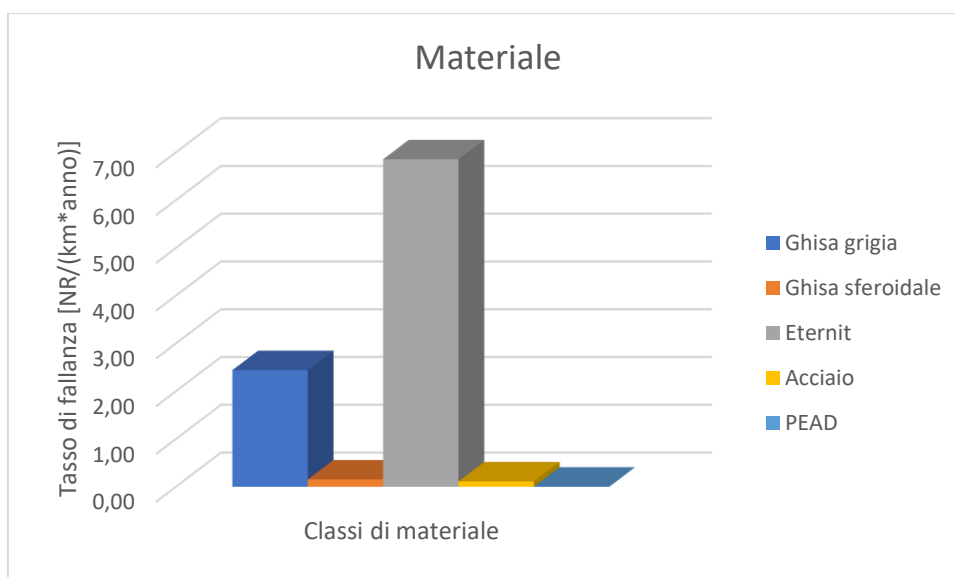


Figura 4-32 Tasso di fallanza dei materiali presenti nella tabella *RegressioneCondotte*

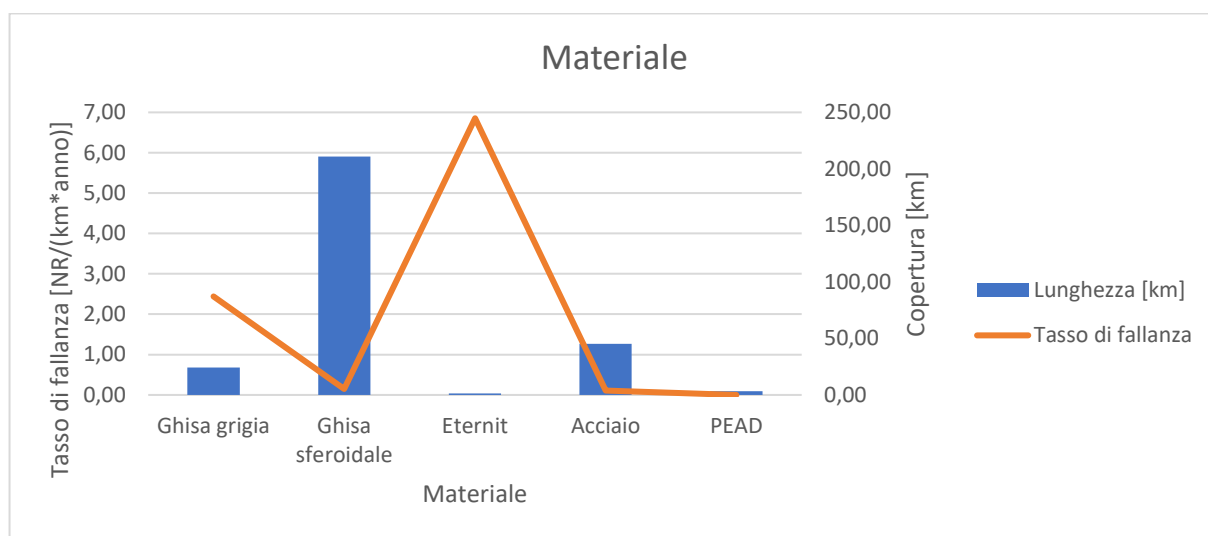


Figura 4-33 Tasso di rottura osservato e copertura offerta dei singoli materiali presenti nella tabella *RegressioneCondotte*

L'eternit e la ghisa grigia risultano i materiali più esposti alle rotture con valori del tasso di fallanza elevati.

Calcolo del tasso di fallanza delle classi di diametro

Allo stesso modo è stato valutato il tasso di fallanza delle tre classi di diametro individuate.

Tabella 4-27 Calcolo del tasso di fallanza delle classi di diametro presenti all'interno della tabella *RegressioneCondotte*

Classe	Diametro [mm]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	D ≤ 100 mm	113.61	39.77	78.00	73.58	0.69
2	100 mm < D < 200 mm	81.29	28.45	25.00	23.58	0.31
3	D ≥ 200 mm	90.79	31.78	3.00	2.83	0.03
Totale		285.69	100.00	106.00	100.00	

Le condotte con diametro minore di 200 millimetri sono le più diffuse. La rappresentazione dei dati su grafico conferma quanto è già stato riscontrato nella tabella *ReteTorino*: i diametri più piccoli vanno incontro a rotture più frequentemente.

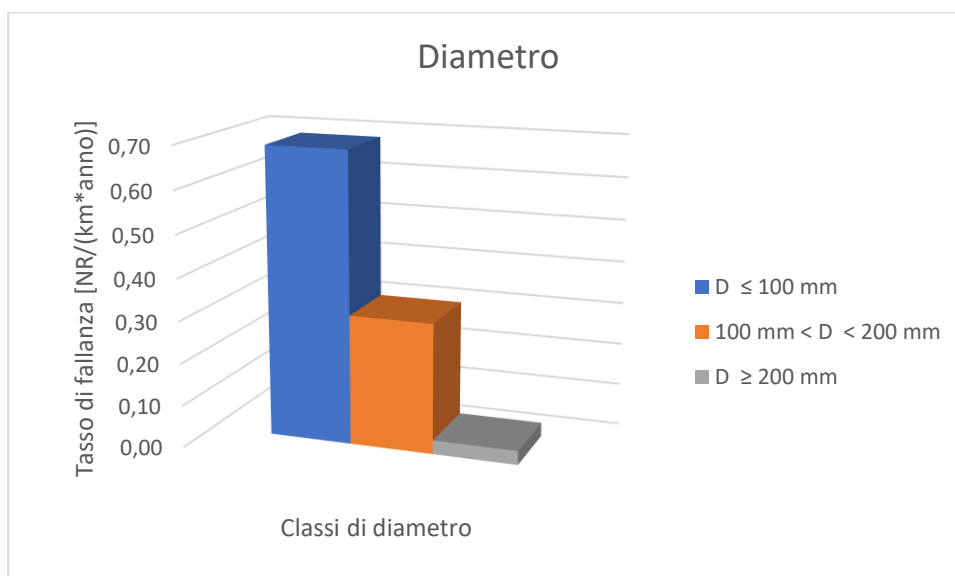


Figura 4-34 Tasso di fallanza delle classi di diametro presenti nella tabella *RegressioneCondotte*

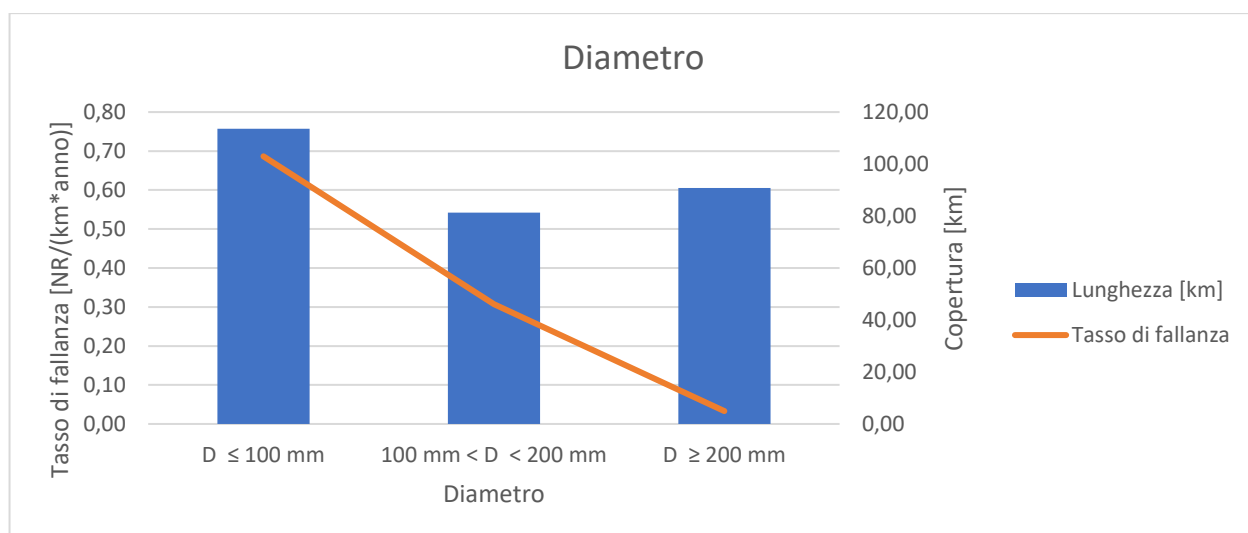


Figura 4-35 Tasso di rottura osservato e copertura offerta delle classi di diametro presenti nella tabella *RegressioneCondotte*

Calcolo del tasso di fallanza delle classi relative all'anno di posa

I risultati ottenuti analizzando il campo anno di posa sono fortemente influenzati dal gran numero di condotte eliminate per ottenere la tabella *RegressioneCondotte*, priva di dati mancanti. Percentuali esigue di condotte, posate in tempi antecedenti il 1990, sono dotate dell'informazione sull'età, a prova che tale dato è stato riportato con maggiore frequenza solo in tempi recenti. Le condotte più vulnerabili sembrano essere quelle posate tra il 1930 e il 1950, anni in cui l'eternit, che si è dimostrato un materiale altamente soggetto a rotture, era fortemente utilizzato.

Tabella 4-28 Calcolo del tasso di fallanza delle classi di anno di posa presenti all'interno della tabella *RegressioneCondotte*

Classe	Anno posa	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	1870 - 1890	0.91	0.32	1.00	0.94	1.10
2	1890 - 1910	9.84	3.44	21.00	19.81	2.13
3	1910 - 1930	7.79	2.73	13.00	12.26	1.67
4	1930 - 1950	2.80	0.98	14.00	13.21	4.99
5	1950 - 1970	5.14	1.80	11.00	10.38	2.14
6	1970 - 1990	31.18	10.91	6.00	5.66	0.19
7	1990 - 2010	212.27	74.30	39.00	36.79	0.18
8	2010 →	15.76	5.52	1.00	0.94	0.06
Totale		285.69	100.00	106.00	100.00	

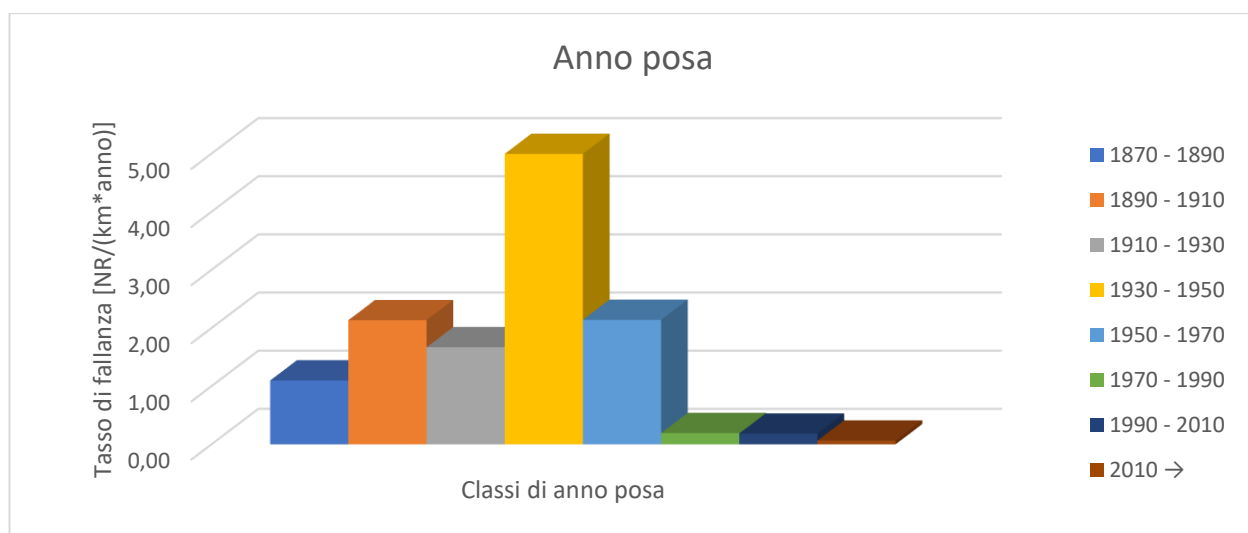


Figura 4-36 Tasso di fallanza delle classi di anno di posa presenti nella tabella RegressioneCondotte

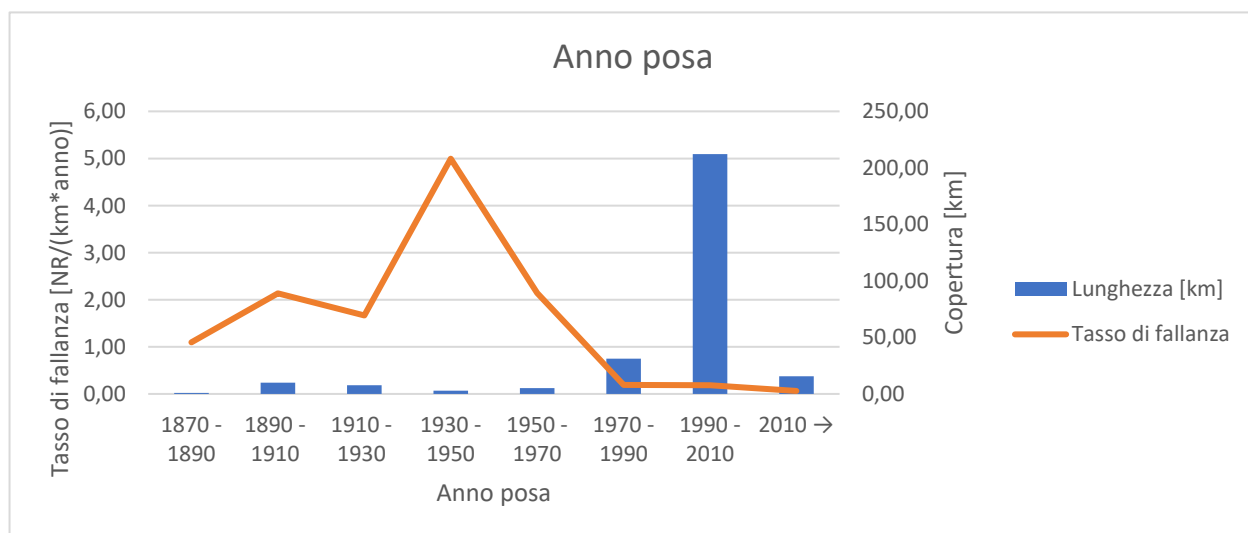


Figura 4-37 Tasso di rottura osservato e copertura offerta delle classi di età presenti nella tabella RegressioneCondotte

Calcolo del tasso di fallanza delle classi di carico massimo

Sono stati analizzati i carichi massimi delle condotte suddivise in cinque differenti classi.

Tabella 4-29 Calcolo del tasso di fallanza delle classi di carico massimo presenti nella tabella RegressioneCondotte

Classe	Carico max [m]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	(0, 40] m	54.54	19.08	10.00	9.43	0.18
2	(40, 80] m	202.98	71.01	88.00	83.02	0.43
3	(80, 120] m	7.76	2.71	2.00	1.89	0.26
4	(120,200] m	12.42	4.34	2.00	1.89	0.16
5	>200 m	8.14	2.85	4.00	3.77	0.49
Totale		285.84	100.00	106.00	100.00	

Circa il 71% delle condotte presenti nella tabella *RegressioneCondotte* è caratterizzata da un carico massimo compreso tra 40 e 80 metri.

Le condotte maggiormente presenti risultano, in seguito al calcolo del tasso di fallanza, anche tra le più vulnerabili, infatti l'83% dei guasti totali contenuti in tabella appartiene alla classe due.

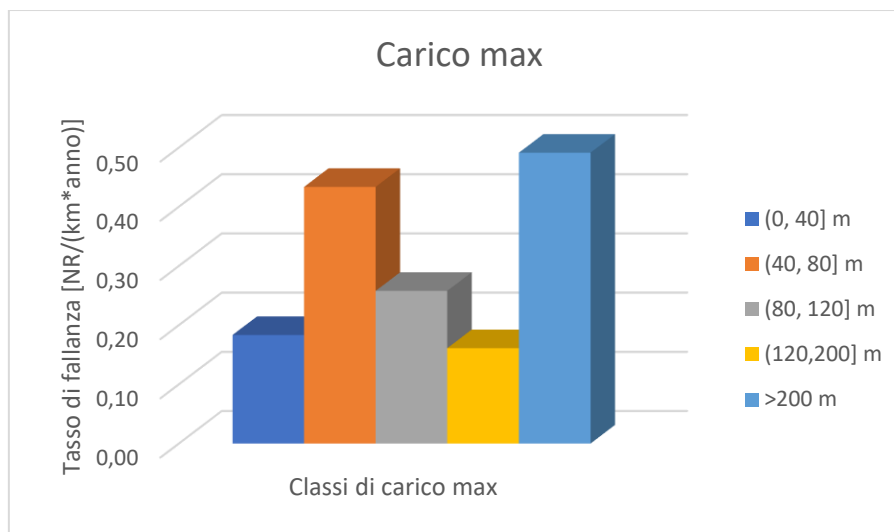


Figura 4-38 Tasso di fallanza delle classi di carico massimo presenti nella tabella *RegressioneCondotte*

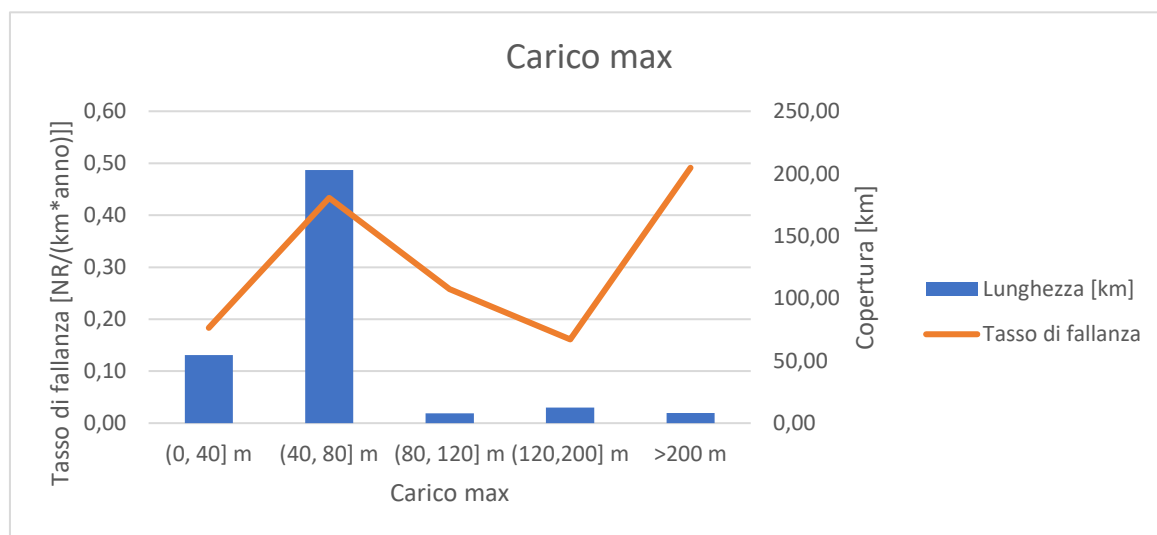


Figura 4-39 Tasso di rottura osservato e copertura delle classi di carico massimo presenti in tabella *RegressioneCondotte*

Calcolo del tasso di fallanza delle classi di variazione di carico

Infine si riporta l'analisi delle classi di variazione di carico in cui sono state suddivise le condotte.

Il tasso di fallanza mostra piccole variazioni rispetto al valor medio, pari a 0.30, in tutte le classi. Le più estese in termini di chilometri sono quelle relative a variazioni di carico non superiori a 20 metri. L'eliminazione di gran parte dei dati durante la costruzione della tabella *RegressioneCondotte*, non consente di poter ben determinare l'andamento delle fluttuazioni nelle classi con valori maggiori di 40 metri, infatti il campione analizzato è poco significativo.

Tabella 4-30 Calcolo del tasso di fallanza delle classi di variazione di carico presenti nella tabella *RegressioneCondotte*

Classe	ΔH [m]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	(0, 10] m	81.11	28.39	18.00	16.98	0.22
2	(10, 20] m	193.68	67.79	85.00	80.19	0.44
3	(20, 30] m	4.47	1.56	1.00	0.94	0.22
4	(30,40] m	3.15	1.10	1.00	0.94	0.32
5	> 40 m	3.27	1.15	1.00	0.94	0.31
Totale		285.69	100.00	106.00	100.00	

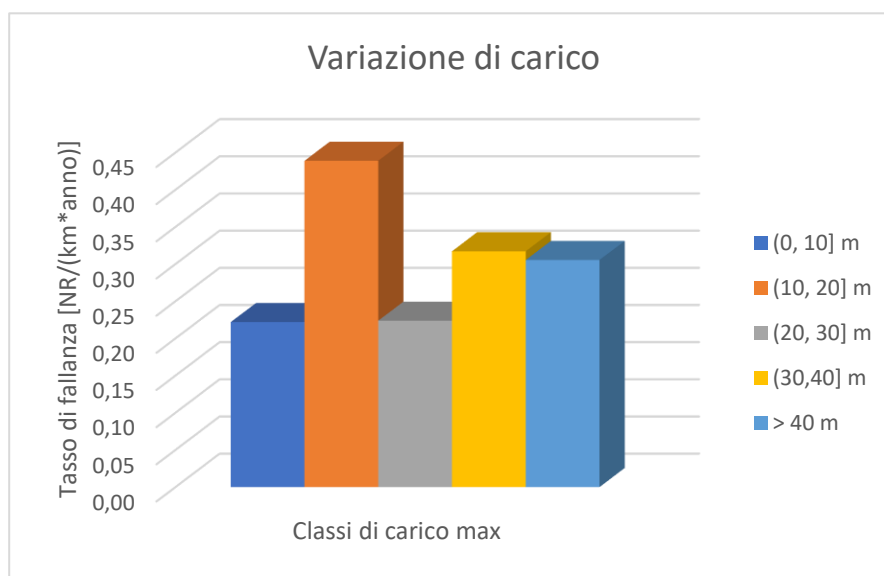


Figura 4-40 Tasso di fallanza delle classi di variazioni di carico presenti nella tabella *RegressioneCondotte*

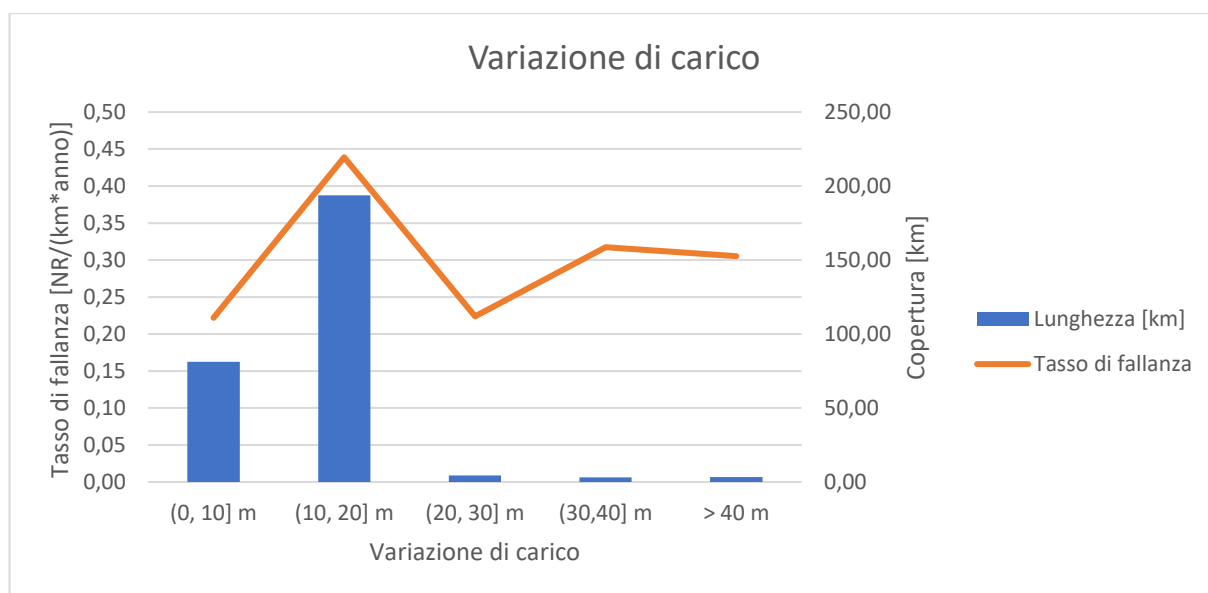


Figura 4-41 Tasso di rottura osservato e copertura delle classi di variazioni di carico presenti in tabella *RegressioneCondotte*

4.3 ANALISI DELLA TABELLA REGRESSIONE TGM

La tabella *RegressioneTGM*, come già descritto, è ottenuta dall'estrazione delle condotte, appartenenti alla tabella *RegressioneCondotte*, che presentano sovrapposizione spaziale con i soli tratti stradali di cui si hanno a disposizione i dati di traffico.

La tabella *RegressioneTGM* contiene fattori di tipo strutturale (anno di posa, diametro, materiale), operativo (pressione massima di esercizio e variazione di carico) e ambientale (flusso veicolare, velocità media del carico dinamico). Questi ultimi rappresentano la vera novità rispetto le tabelle precedentemente introdotte e analizzate.

La tabella *RegressioneTGM* contiene 63 rotture, difatti la tabella è frutto di una scrematura fatta per soddisfare l'esigenza di disporre di un quadro contenente solo le condotte avente tutte e otto le variabili indipendenti.

Siccome la regressione logistica si baserà sulla tabella *RegressioneTGM*, si procede alla descrizione della sua composizione.

Calcolo del tasso di fallanza delle classi di materiale

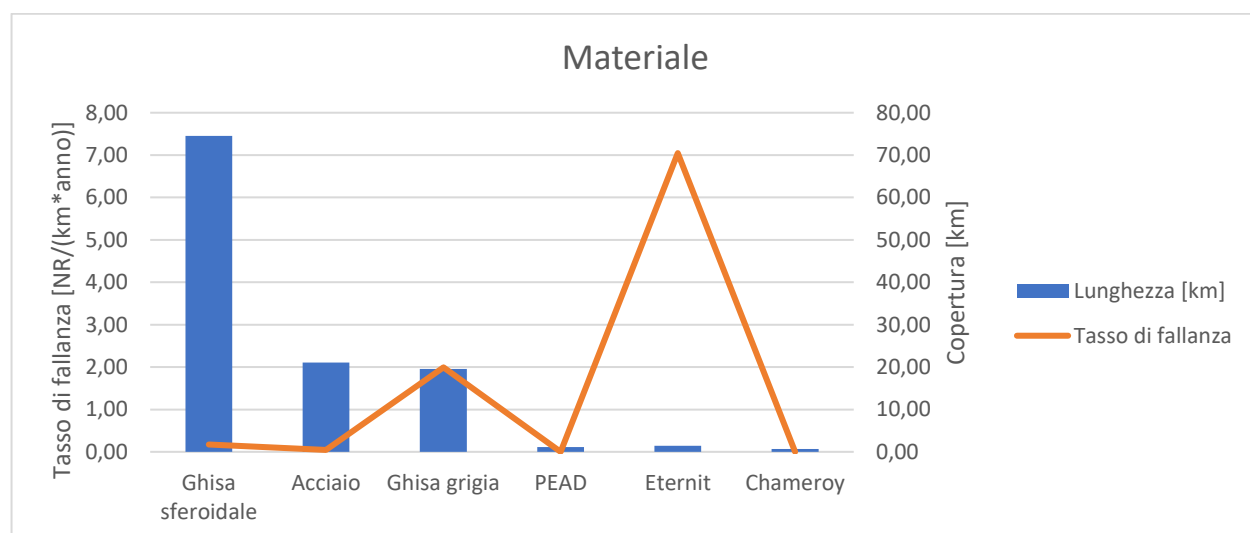
La variabile materiale risulta, anche in questo paragrafo, la prima ad essere analizzata.

Le tipologie di materiale indagabili sono state ridotte a sei, ossia quei materiali che caratterizzano la rete di Torino per almeno lo 0.5% della sua estensione.

Tabella 4-31 Calcolo del tasso di fallanza dei singoli materiali presenti all'interno della tabella *RegressioneTGM*

Materiale	N° materiale	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
Ghisa sferoidale	4	74.53	62.95	13	20.63	0.17
Acciaio	7	21.09	17.81	1	1.59	0.05
Ghisa grigia	2	19.53	16.50	39	61.90	2.00
PEAD	6	1.11	0.93	0	0.00	0.00
Eternit	1	1.42	1.20	10	15.87	7.05
Chameroy	5	0.71	0.60	0	0.00	0.00
Totale		118.38	99.98	63	100.00	

L'analisi mostra che i materiali che presentano maggiore tendenza alla rottura risultano l'eternit e la ghisa grigia. Si osserva che i valori mostrati per ciascun materiale non sono differenti da quelli forniti dall'analisi della tabella *RegressioneCondotte*.

Figura 4-42 Tasso di fallanza dei materiali presenti nella tabella *RegressioneTGM*

Calcolo del tasso di fallanza delle classi di diametro

Applicando gli stessi passaggi è stato valutato il tasso di fallanza delle tre classi di diametro in cui sono state suddivise le condotte.

Tabella 4-32 Calcolo del tasso di fallanza delle classi di diametro presenti all'interno della tabella *RegressioneTGM*

Classe	Diametro [mm]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	D ≤ 100 mm	27.99	23.63	42	66.67	1.50
2	100 mm < D < 200 mm	36.96	31.21	19	30.16	0.51
3	D ≥ 200 mm	53.47	45.16	2	3.17	0.04
Totale		118.414	100	63	100	

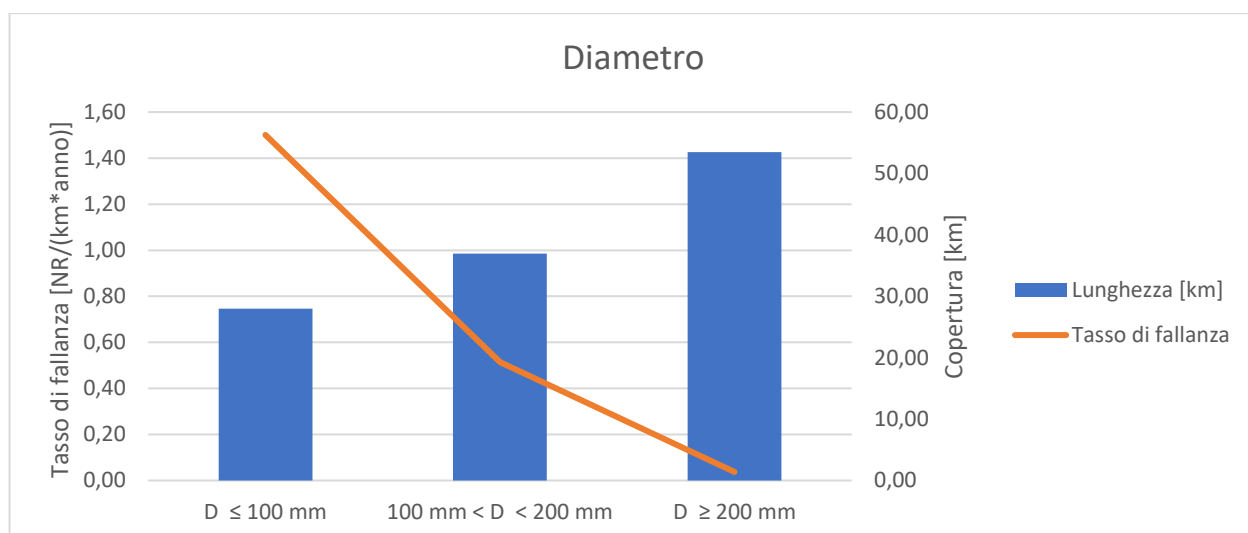


Figura 4-43 Tasso di rottura osservato e copertura offerta delle classi di diametro presenti nella tabella *RegressioneTGM*

Si evidenzia, come i risultati precedenti lasciano prevedere, un andamento decrescente del tasso di rottura con l'aumentare del diametro.

Calcolo del tasso di fallanza delle classi relative all'anno di posa

Le condotte sono state suddivise in classi ventennali ed è stato calcolato il tasso di fallanza.

Tabella 4-33 Calcolo del tasso di fallanza delle classi di anno di posa presenti all'interno della tabella *RegressioneTGM*

Classe	Anno posa	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	1870 - 1890	0.61	0.51	1	1.59	1.64
2	1890 - 1910	8.80	7.43	18	28.57	2.04
3	1910 - 1930	6.58	5.55	10	15.87	1.52
4	1930 - 1950	2.46	2.07	14	22.22	5.70
5	1950 - 1970	1.29	1.09	0	0.00	0.00
6	1970 - 1990	14.96	12.64	1	1.59	0.07
7	1990 - 2010	79.04	66.75	19	30.16	0.24
8	2010 →	4.68	3.95	0	0.00	0.00
Totale		118.414	100	63	100	

Circa il 67% delle condotte presenti in tabella sono state posate nel ventennio 1990-2010. I tassi di rottura più elevati si registrano nelle prime quattro classi con un picco proprio in corrispondenza della quarta classe.

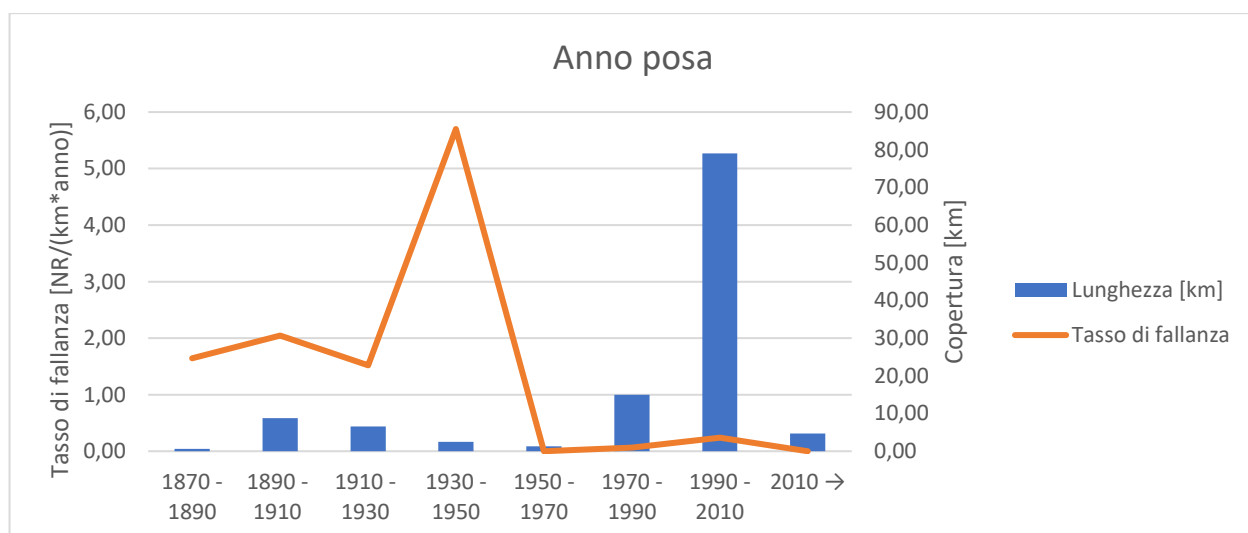


Figura 4-44 Tasso di rottura osservato e copertura offerta delle classi di anno di posa presenti nella tabella
RegressioneTGM

Calcolo del tasso di fallanza delle classi di carico massimo

Anche nel caso del carico massimo è stato valutato il tasso di fallanza per cinque differenti classi.

Tabella 4-34 Calcolo del tasso di fallanza delle classi di carico massimo presenti all'interno della tabella
RegressioneTGM

Classe	Carico max [m]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	(0, 40] m	24.569	20.75	4	6.35	0.16
2	(40, 80] m	91.646	77.39	58	92.06	0.63
3	(80, 120] m	0.406	0.34	0	0.00	0.00
4	(120,200] m	0.979	0.83	1	1.59	1.02
5	>200 m	0.814	0.69	0	0.00	0.00
Totale		118.414	100.00	63	100.00	

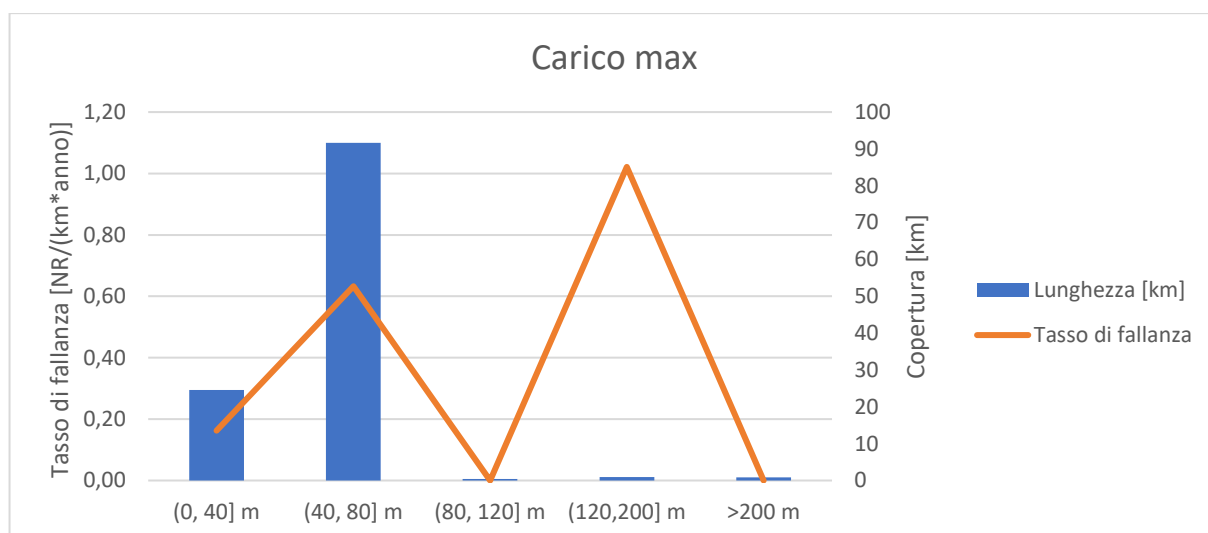


Figura 4-45 Tasso di rottura osservato e copertura offerta delle classi di carico massimo presenti nella tabella *RegressioneTGM*

La maggior parte dei tratti è caratterizzata da un carico massimo compreso tra 40 e 80 metri; tali condotte risultano caratterizzate da un valore di tasso di fallanza elevato, seppur le più vulnerabili siano quelle con carichi compresi fra 120 e 200 metri.

Calcolo del tasso di fallanza delle classi di variazione di carico

Si riportano in Tabella 4-35 i risultati ottenuti classificando le condotte in cinque classi.

Tabella 4-35 Calcolo del tasso di fallanza delle classi di variazione di carico presenti all'interno della tabella *RegressioneTGM*

Classe	ΔH [m]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	(0, 10] m	22.40	18.92	2.00	3.17	0.09
2	(10, 20] m	94.00	79.40	60.00	95.24	0.64
3	(20, 30] m	1.03	0.87	1.00	1.59	0.97
4	(30,40] m	0.50	0.42	0.00	0.00	0.00
5	> 40 m	0.46	0.39	0.00	0.00	0.00
Totale		118.38	100.00	63.00	100.00	

La quasi totalità delle condotte presenta variazione di carico compresa tra 10 e 20 metri, a tale classe corrisponde anche il 95% dei guasti contenuti nella tabella *RegressioneTGM*. L'assenza di dati, dovuta alle operazioni di filtraggio, non consente di stimare il tasso di fallanza per valori maggiori.

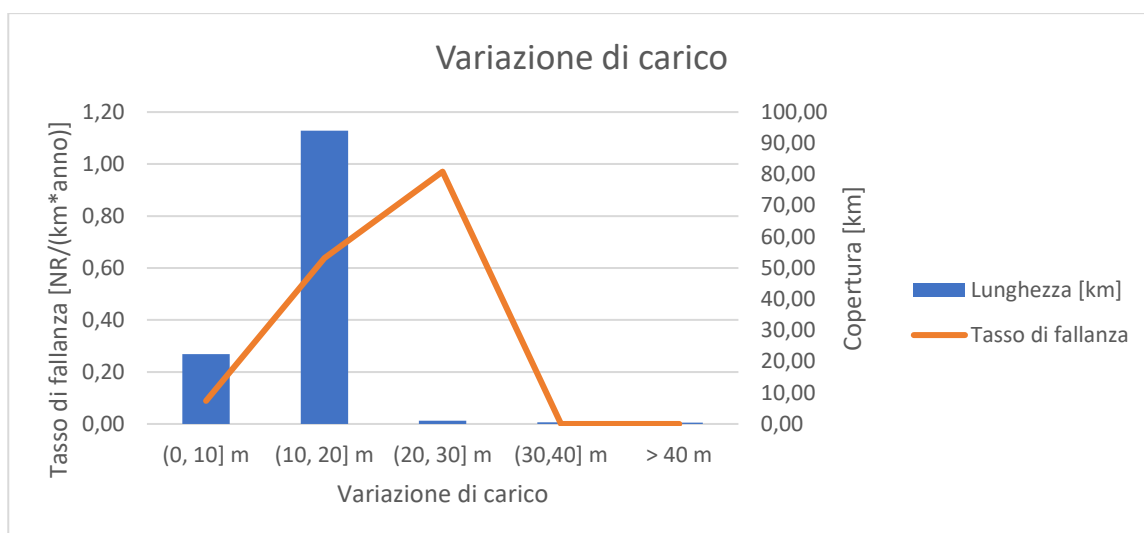


Figura 4-46 Tasso di rottura osservato e copertura delle classi di variazione di carico presenti in tabella *RegressioneTGM*

Calcolo del tasso di fallanza delle classi di velocità media

Sono state individuate cinque classi di velocità media e per ogni classe è stata valutata la lunghezza in chilometri, il numero di guasti e applicando l'equazione (1), è stato calcolato il tasso di fallanza. I dati di traffico forniti dalla società *5T* risultano ben distribuiti sulle arterie principali di Torino, quali i grandi viali, e risultano molto dettagliati nella zona centrale della città in cui la velocità di percorrenza è limitata. Per questo motivo, vi è un'elevata percentuale di condotte associate a strade con velocità minori di 40 km/h.

Tabella 4-36 Calcolo del tasso di fallanza delle classi di velocità media presenti all'interno della tabella *RegressioneTGM*

Classe	Velocità [km/h]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	(0, 20]	35.74	30.18	35	55.56	0.98
2	(20, 30]	55.40	46.78	27	42.86	0.49
3	(30, 40]	16.74	14.14	1	1.59	0.06
4	(40, 80]	9.92	8.37	0	0.00	0.00
5	(80, 150]	0.62	0.52	0	0.00	0.00
Totale		118.414	100	63	100	

Si osserva che il tasso di fallanza decresce all'aumentare della velocità.

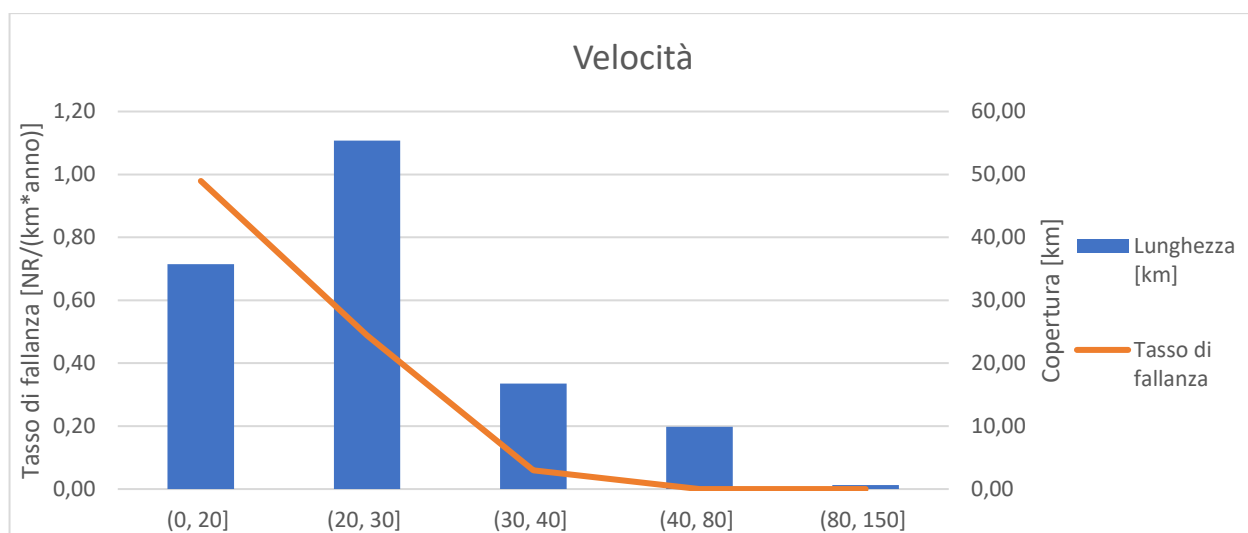


Figura 4-47 Tasso di rottura osservato e copertura delle classi di velocità media presenti in tabella *RegressioneTGM*

Calcolo del tasso di fallanza delle classi di flusso veicolare

È stato valutato il tasso di fallanza di sette diverse classi di flusso veicolare, misurato in veicolo/ora.

Tabella 4-37 Calcolo del tasso di fallanza delle classi di flusso veicolare presenti all'interno della tabella *RegressioneTGM*

Classe	Flusso [veic/h]	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
1	(0, 200]	12.887	10.88	8	12.70	0.62
2	(200, 400]	33.574	28.35	16	25.40	0.48
3	(400, 600]	33.923	28.65	30	47.62	0.88
4	(600, 800]	15.048	12.71	3	4.76	0.20
5	(800,1200]	16.763	14.16	6	9.52	0.36
6	(1200,2000]	5.867	4.95	0	0.00	0.00
7	(2000,4000]	0.352	0.30	0	0.00	0.00
Totale		118.414	100.00	63	100	

Una elevata incidenza del carico da traffico si ha nel centro cittadino. Il tasso di fallanza, in linea generale, decresce all'aumentare del flusso veicolare.

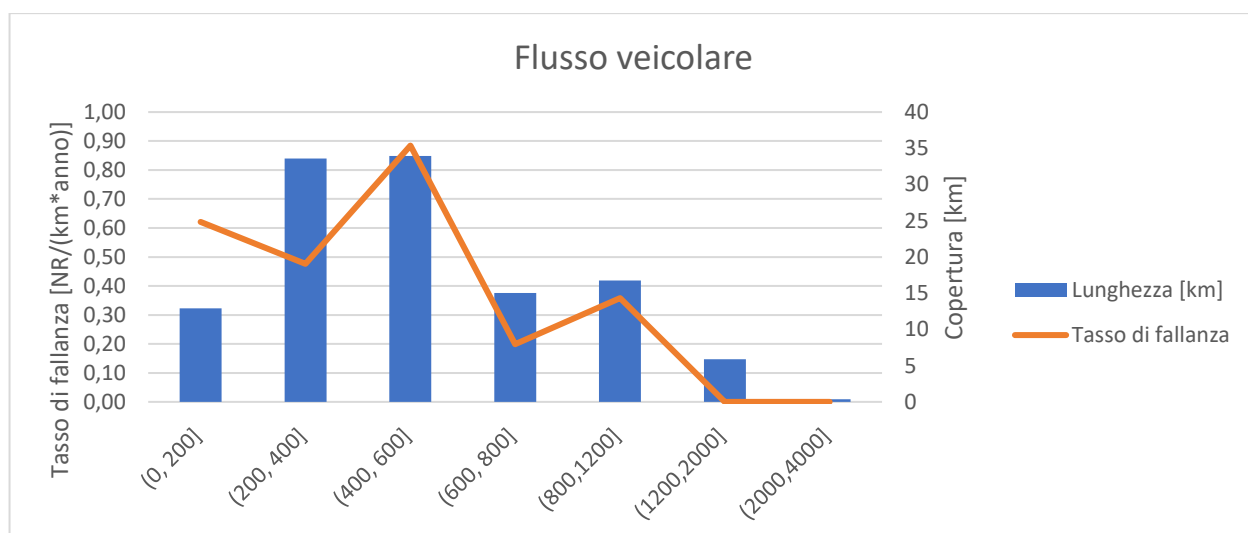


Figura 4-48 Tasso di rottura osservato e copertura delle classi di flusso veicolare presenti in tabella *RegressioneTGM*

Calcolo del tasso di fallanza delle condotte soggette o non al passaggio del tram

Infine, si riportano i dati relativi alla classificazione delle condotte in funzione della presenza (1) o assenza (0) del tram.

Tabella 4-38 Calcolo del tasso di fallanza delle condotte soggette (1) o non soggette (0) alle vibrazioni indotte dal tram

Tram route	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
0	90.189	76.16	31	49.21	0.34
1	28.225	23.84	32	50.79	1.13
Totale	118.414	100	63	100	

Il 76% delle condotte presenti nella tabella *RegressioneTGM* non è spazialmente sovrapposto a strade munite di binari per il passaggio dei tram. Il tasso di fallanza massimo lo si riscontra, invece, nella classe di condotte la cui ubicazione coincide con il percorso effettuato dai tram.

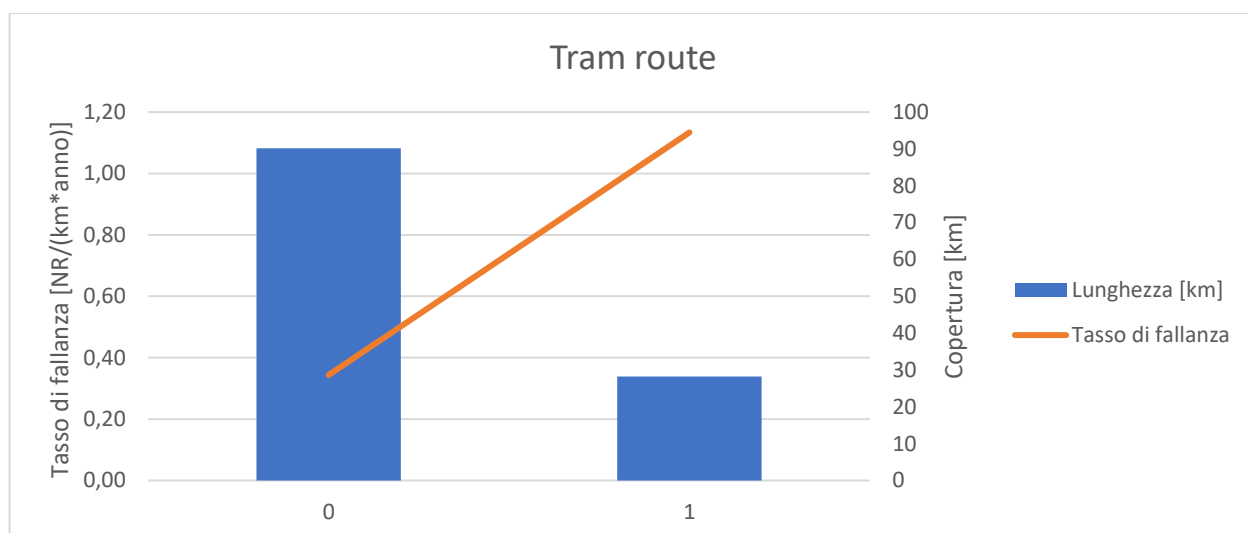


Figura 4-49 Tasso di rottura osservato e copertura delle classi di condotte soggette (1) o non (0) al passaggio del tram

4.3.1 Analisi combinata della variabile 'materiale' e 'anno di posa'

L'andamento dei valori del tasso di fallanza calcolato, mostra per la variabile 'anno di posa' nelle tre tabelle analizzate, un picco in corrispondenza della classe ventennale 1930-1950. Il valore massimo comune trova giustificazione nella ricerca dei tecnici e dei progettisti di condotte con caratteristiche, quali: la facilità di posa, l'economicità, la resistenza ai carichi, la lunga durata, la resistenza alla corrosione via via migliori, che si rispecchia nell'evoluzione dell'industria dei materiali.

Si riporta nel grafico in Figura 4-50 l'evoluzione dell'utilizzo dei materiali in rete.

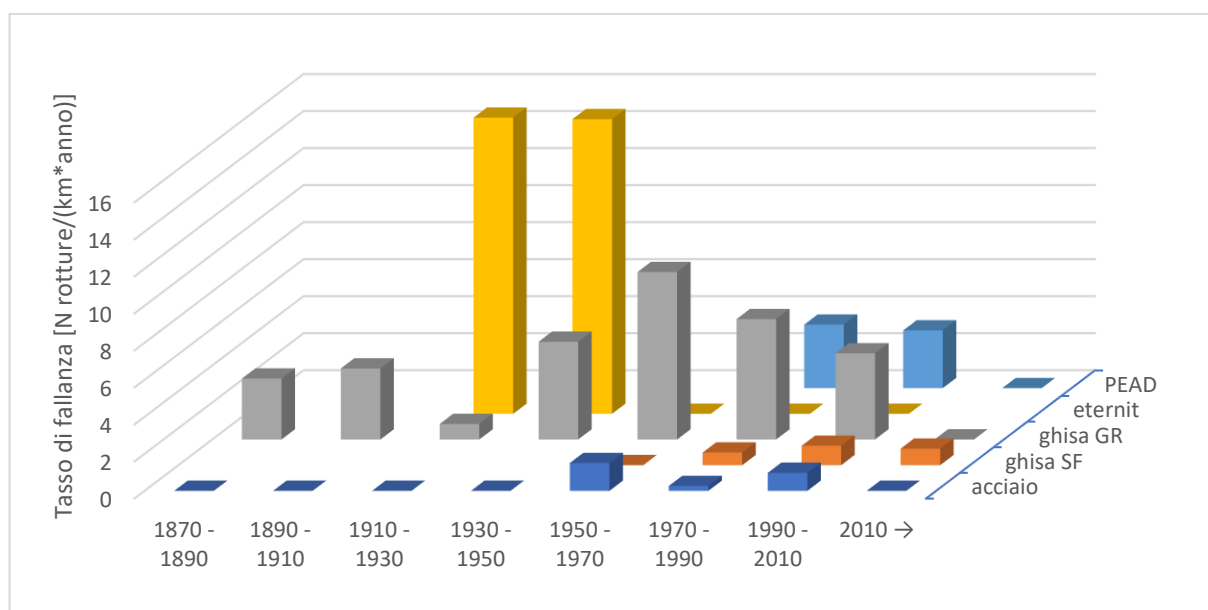


Figura 4-50 Tasso di rottura dei materiali più presenti in rete in funzione dell'anno di posa

Il picco è motivato dall'introduzione sul mercato delle condotte in eternit, largamente utilizzate in passato ed ancora presenti in rete, che hanno manifestato un numero di rotture elevato, se rapportato alla loro estensione. Si nota che materiali, come la ghisa sferoidale ed il PEAD, popolano il grafico solo a partire dagli anni '50, perché utilizzate da tempi recenti.

4.4 CORRELAZIONI TRA LE VARIABILI INDIPENDENTI

Dalla selezione delle variabili da utilizzare nell'analisi scaturisce l'esito finale di ogni tecnica multivariata, e quindi della regressione multipla. Poiché la regressione multipla è una metodologia che studia relazioni di dipendenza, è necessario specificare qual è la variabile dipendente e quali sono le variabili indipendenti. L'ultima è quella che dovrebbe essere in grado di influenzare l'altra. È importante segnalare che i dati presentati nel database possono essere di due tipi in termini di correlazione con altre variabili:

- non correlati a nessun parametro;
- correlati ad uno o più parametri.

Per correlazione tra variabili si intende l'esistenza di una relazione che associ ad ogni valore della prima (X) un determinato valore della seconda (Y). Quindi, due variabili, per esser definite correlate, devono variare una in funzione dell'altra, positivamente o negativamente. La correlazione viene stimata attraverso il concetto di covarianza, in particolare il rapporto tra la covarianza osservata e la covarianza massima possibile definisce il coefficiente di correlazione r . La sua formulazione è:

$$r_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \quad (2)$$

Il coefficiente indica l'entità della correlazione e il suo valore varia tra -1 ed 1. I valori estremi indicano una perfetta correlazione negativa o positiva, mentre il valore 0 indica totale assenza di correlazione. Orientativamente una correlazione viene definita:

- alta se vi è una relazione di associazione marcata e il coefficiente r è compreso tra $\pm 0,71$ - $0,90$;
- media, con relazione di associazione sostantiva, se il coefficiente r è compreso tra $\pm 0,41$ - $0,70$;
- bassa se il coefficiente r è compreso tra $\pm 0,21$ - $0,40$;

Per valutare la correlazione tra le variabili presenti nella tabella *RegressioneTGM* è stato calcolato il coefficiente di correlazione. Si riportano i risultati in Tabella 4-39.

Tabella 4-39 Correlazione fra le variabili

	Num materia le	diametr o	anno posa	pMax	pRange	flag	N. Rotture	velocità	flusso veicol	Tram route
Num materia le	1.000	0.485	0.541	0.019	0.008	-0.147	-0.144	0.203	0.212	-0.008
diametr o	0.485	1.000	0.138	-0.077	-0.054	-0.079	-0.070	0.176	0.281	0.067
anno posa	0.541	0.138	1.000	0.026	0.018	-0.164	-0.152	0.195	0.148	-0.165
pMax	0.019	-0.077	0.026	1.000	0.736	-0.001	-0.001	0.220	-0.062	-0.019
pRange	0.008	-0.054	0.018	0.736	1.000	-0.008	-0.007	0.173	-0.022	-0.024
flag	-0.147	-0.079	-0.164	-0.001	-0.008	1.000	0.860	-0.047	-0.036	0.061
N. Rotture	-0.144	-0.070	-0.152	-0.001	-0.007	0.860	1.000	-0.044	-0.021	0.075
velocità	0.203	0.176	0.195	0.220	0.173	-0.047	-0.044	1.000	0.573	-0.137
flusso veicol	0.212	0.281	0.148	-0.062	-0.022	-0.036	-0.021	0.573	1.000	0.123
Tram route	-0.008	0.067	-0.165	-0.019	-0.024	0.061	0.075	-0.137	0.123	1.000

Si nota che esiste correlazione bassa o molto bassa tra le variabili e degna di nota è la correlazione media tra le coppie:

- materiale – anno di posa, $r=0.541$;

- materiale – diametro, $r=0.485$;
- velocità – flusso veicolare, $r=0.573$.

È possibile parlare di correlazione forte tra la coppia:

- pressione massima – variazione di carico, $r=0.736$.

Motivo per cui l'analisi di regressione non si baserà sulla variabile '*variazione di carico*' in quanto marcatamente associata alla pressione massima.

4.5 CONCLUSIONI

In conclusione, l'analisi dei dati, contenuti nella tabella *ReteTorino*, *RegressioneCondotte*, *RegressioneTGM*, ha previsto il calcolo delle occorrenze e il calcolo del tasso di fallanza delle diverse classi in cui sono state suddivise le variabili. I risultati ottenuti consentono di affermare che:

- i materiali più vulnerabili, che presentano tasso di fallanza elevato, sono l'eternit e la ghisa grigia, che risulta il materiale più presente in rete;
- il tasso di fallanza diminuisce all'aumentare del diametro;
- l'informazione relativa all'anno di posa delle condotte è spesso mancante e le indicazioni fornite dalla sua analisi variano da caso a caso. Il valore di picco del tasso di fallanza è però comune nelle tre tabelle: le condotte posate nel ventennio 1930-1950 sono le più vulnerabili;
- gran parte delle condotte sono ubicate in corrispondenza di strade percorribili a media-bassa velocità e caratterizzate da flussi veicolari ridotti, a cui sembrerebbero corrispondere i tassi di fallanza più alti. È bene evidenziare però che le classi rappresentanti valori di velocità e di flusso veicolare medio-alti potrebbero restituire un valore falsato a causa della ridotta significatività del campione, dovuta alle tante operazioni di filtraggio dati. Tra le possibili ragioni del valore contenuto del tasso di fallanza di condotte posate lungo strade molto trafficate e percorse ad elevate velocità, potrebbero esserci l'utilizzo di accortezze maggiori nella fase di posa;
- le condotte posate in strade soggette al passaggio del tram sono le più vulnerabili.

Infine il calcolo dei coefficienti di correlazione segnala relazione di associazione tra le variabili '*pressione massima*' e '*variazione di carico*', di conseguenza nello sviluppo della regressione logistica, sarà necessario prendere in considerazione solo un campo fra i due.

5 LA REGRESSIONE LOGISTICA

L'utilizzo di un modello soddisfa l'obiettivo di dimostrare che una variabile di un set di dati costituisca un fattore di rischio non identificato in precedenza. Ciò sottintende la necessità di comprovare che la presenza di questo fattore di rischio effettivamente si associ con l'evento studiato. I modelli di regressione multivariati consentono di calcolare una misura dell'associazione fra fattore di rischio e l'evento, e studiare l'associazione tenendo conto delle altre caratteristiche. Come anticipato, per modellare le rotture della rete di distribuzione idrica del comune di Torino si è scelto di utilizzare il modello di regressione logistica.

5.1 IL MODELLO DI REGRESSIONE LOGISTICA

Il modello logistico appartiene all'insieme dei modelli di regressione caratterizzati da una variabile dipendente binaria. Questa, definita anche outcome, può assumere 2 valori: ad una risposta negativa (evento assente) è assegnato il valore 0, ad un outcome positivo (evento presente) è assegnato 1. Il modello logistico ha la vantaggiosa capacità di permettere di esplorare come ogni variabile indipendente, sia di tipo qualitativo che quantitativo, influenzi la probabilità di avvenimento del caso studio. Definite con x_n le variabili indipendenti, che rappresentano le caratteristiche degli elementi che costituiscono l'osservazione campionaria e con Y la variabile binaria, la regressione logistica, anziché modellare direttamente Y , propone un modello per la probabilità che Y appartenga ad una delle due categorie. L'associazione fra la probabilità dell'evento e le variabili indipendenti nella regressione logistica non è lineare. La probabilità che l'evento si verifichi, sempre compresa nell'intervallo $[0, 1]$, è data da:

$$p(Y = 1|x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)} \quad (3)$$

Il parametro β è definito coefficiente di regressione e rappresenta il cambiamento in Y per ogni variazione unitaria della x .

5.1.1 La stima dei parametri

I coefficienti dell'algoritmo vengono stimati tramite il metodo della massima verosimiglianza. Questo metodo prevede di ricavare la serie di coefficienti di regressione per i quali la probabilità di ottenere i dati osservati è massima; tale implementazione

avviene usando un efficiente algoritmo di ottimizzazione numerica, che prevede un processo iterativo, partendo da valori arbitrari dei parametri. Supponendo di considerare un modello con una sola variabile indipendente x , il procedimento che consente di ottenere i parametri β_0 e β_1 è il seguente:

$$L(\beta_0, \beta_1) = \prod_{i=1}^n f(y_i, x_i) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (4)$$

Si ricava ora la funzione di log-verosimiglianza:

$$\ln L(\beta_0, \beta_1) = \sum_{i=1}^n \{y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)]\} \quad (5)$$

Con una serie di passaggi intermedi si arriva a:

$$l(\beta_0, \beta_1) = \sum_{i=1}^n \{y_i(\beta_0 + \beta_1 x_i) - \ln[1 + e^{\beta_0 + \beta_1 x_i}]\} \quad (6)$$

Calcolando le derivate parziali della (6) rispetto ai parametri β_0 e β_1 e ponendole uguali a 0, si ricava il sistema di equazioni di verosimiglianza, la cui soluzione restituisce le stime di massima verosimiglianza, indicate rispettivamente con b_0 e b_1 :

$$\sum_{i=1}^n \left\{ y_i - \frac{1}{[1 + e^{b_0 + b_1 x_i}]} e^{b_0 + b_1 x_i} \right\} = 0 \quad (7)$$

$$\sum_{i=1}^n \left\{ y_i x_i - \frac{1}{[1 + e^{b_0 + b_1 x_i}]} e^{b_0 + b_1 x_i} x_i \right\} = 0 \quad (8)$$

La soluzione della (7) e della (8) coincide con l'implementazione delle stime di massima verosimiglianza e avviene generalmente utilizzando software quali: SAS PROC GENMOD, JMP, Gretl e Minitab.

5.1.2 Funzione Logistica

La regressione logistica si avvale della funzione logistica per determinare la classificazione dei valori di input. La funzione logistica, che su grafico è una curva descritta da un arcsinusoide, converte un numero di valore reale in numeri compresi tra 0 e 1, estremi inclusi. Maggiore è il valore reale, maggiore sarà vicino a 1 il risultato e viceversa. L'equazione che descrive la funzione logistica è:

$$F(x) = \frac{1}{1 + \exp^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}} \quad (9)$$

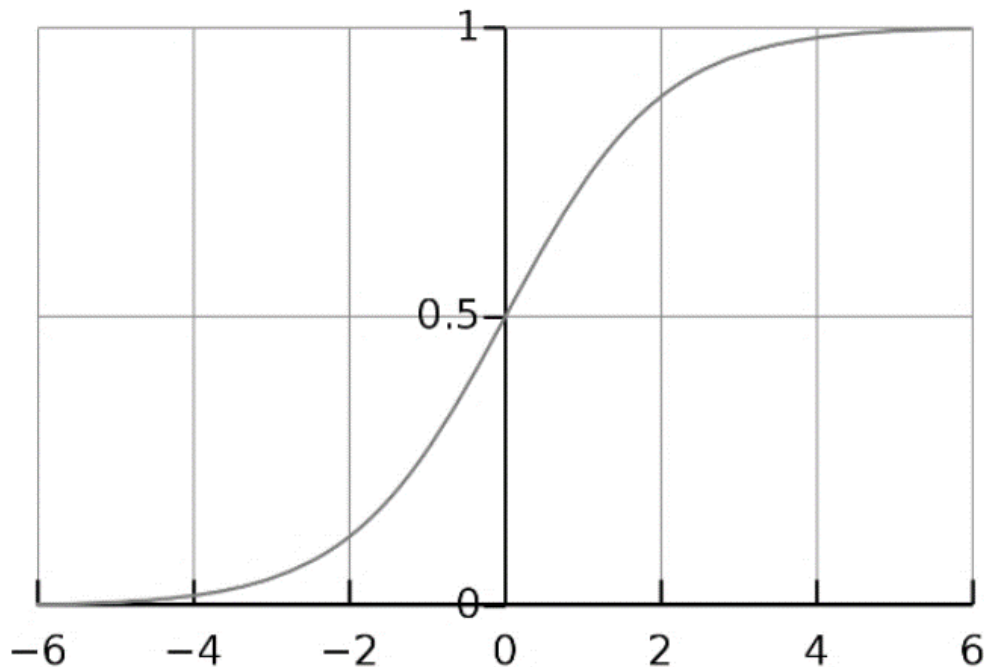


Figura 5-1 Grafico della funzione logistica

5.1.3 Logit e Odds

Per ottenere l'equazione di regressione logistica espressa in termini probabilistici è necessario eseguire una serie di accorgimenti matematici, inserendo nell'equazione (3) le probabilità. Applicando le proprietà che legano esponenziali e logaritmi, la regressione logistica può essere espressa come:

$$\ln\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (10)$$

Dove:

- ❖ $\ln\left(\frac{p(x)}{1-p(x)}\right)$ è definita funzione *logit* e permette di associare le probabilità (quindi un valore compreso tra 0 e 1) all'intero intervallo dei numeri reali. Questa rappresenta una funzione di collegamento ed è l'inversa della funzione logistica. L'associazione fra il *logit* e le variabili esplicative è lineare, contrariamente alla formulazione precedente, ed è più facile da modellare;
- ❖ $\left(\frac{p(x)}{1-p(x)}\right)$ in statistica è definito *Odds* e rappresenta il rapporto di probabilità $p(x)$ e la probabilità di non evento $(1-p(x))$. L'*Odds* può assumere qualsiasi valore in $[0, \infty)$ e in particolare:
 - valori pari a 1 indicano che il successo e l'insuccesso sono equiprobabili;
 - valori positivi minori di 1 indicano che il successo è meno probabile dell'insuccesso;
 - valori maggiori di 1 indicano che il successo è più probabile dell'insuccesso.

Manipolando la (10), l'*Odds* può essere espresso anche come:

$$Odds = \exp^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n} \quad (11)$$

Un altro parametro da definire è l'*Odds Ratio*, o rapporto fra gli *Odds*, e rispecchia l'associazione tra due variabili, permettendo di definire la relazione tra due categorie in funzione di una terza:

$$Odds\ Ratio = \frac{\frac{p_1(x)}{1-p_1(x)}}{\frac{p_2(x)}{1-p_2(x)}} \quad (12)$$

5.2 COSTRUZIONE DI UN MODELLO DI REGRESSIONE

La scelta delle variabili da includere nel modello si basa sull'obiettivo della ricerca, ma deve essere pilotata dalla padronanza del problema esaminato. L'efficienza del modello è ridotta in presenza di collinearità, che prevede che due variabili siano fra loro correlate, e in presenza di un eccessivo numero di variabili, che causa il fenomeno di overfitting. Nel primo caso le variabili indipendenti vengono ridotte scegliendo una delle due correlate o alternativamente inserire una loro combinazione; nel secondo si può introdurre una regola concreta che limiti il rapporto fra le variabili e i casi a 1:10. Di fondamentale importanza è

la problematica dei dati mancanti: in un modello multivariato, è sufficiente che una delle variabili sia priva di un'informazione perché il caso venga eliminato dall'analisi; la conseguenza è la perdita di potenza dello studio, e diviene importante se i dati mancanti sono numerosi.

Pur avendo a disposizione molteplici predittori è indispensabile pervenire a un modello in grado di descrivere in modo schematico, ma approfondito l'effetto delle variabili in studio. Parodi osserva che “la selezione del modello può essere effettuata attraverso diversi metodi:

- *backward*, inserisce nel calcolatore tutte le variabili del modello e quindi le elimina una alla volta in base alla significatività dei rispettivi coefficienti, ottenuta tramite il *Likelihood Ratio* test. Questa procedura è di tipo ricorsivo e si arresta allorché siano rimaste nel modello esclusivamente le variabili con coefficiente statisticamente significativo. Il metodo tende a sovrastimare il fitting del modello, ma risulta adatto per l'analisi con dati numerosi;
- *forward*, consiste nell'inserire le variabili nel modello una alla volta sulla base della loro significatività statistica. Questo metodo è adatto quando il numero di predittori è elevato rispetto al numero delle osservazioni;
- *stepwise*, consiste nell'inserire e rimuovere in modo ricorsivo le variabili dal modello sulla base della loro significatività statistica;
- “*best model*”, consiste nel fittare tutti i modelli possibili e nello scegliere quello con la migliore bontà di adattamento. Tale procedimento genera una sovrastima del fitting del modello, generando associazioni fasulle.”¹³

5.3 TEST PER LA BONTÀ DI ADATTAMENTO DEL MODELLO LOGISTICO¹⁴

Una volta vagliato un modello che rappresenti i dati sulla base delle ipotesi formulate nel disegno dello studio, occorre confermare la bontà di adattamento dello stesso e valutare la significatività statistica dei parametri stimati, ovvero fornire delle misure quantitative della capacità del modello di interpretare i dati osservati. In particolare, per valutare la bontà di adattamento ai dati di un modello è possibile utilizzare:

- ❖ il test rapporto di verosimiglianza;
- ❖ lo pseudo- R^2 ;
- ❖ test della bontà di adattamento: Devianza, χ^2 di Pearson e test di Hosmer-Lemeshow.

¹³ Parodi, S. (2007). *Applicazione del modello di regressione di Poisson e del modello logistico in Epidemiologia*.

¹⁴ D.C., E., & G.G. (2012). *Introduction to Linear Regression Analysis*, ch 13.2.4.

5.3.1 Il Test Rapporto di Verosimiglianza

Il rapporto di verosimiglianza, indicato con G , si definisce come:

$$G = -2 \log \frac{L(0)}{L(\beta)} \quad (13)$$

Dove:

- $L(0)$ è la massima verosimiglianza in corrispondenza del modello con la sola intercetta;
- $L(\beta)$ è la funzione di massima verosimiglianza in corrispondenza del modello completo.

G è una statistica che si basa sul rapporto tra due verosimiglianze ed segue una distribuzione del χ^2 con un numero di gradi di libertà pari alla differenza tra il numero di parametri del modello completo e quelli del modello ridotto. Tale test prevede di valutare le informazioni che le variabili considerate aggiungono al modello, infatti se le variabili aggiungono molta informazione al modello con la sola intercetta, la verosimiglianza relativa al modello completo $L(\beta)$ sarà molto maggiore di quella che si ottiene considerando il modello con la sola intercetta $L(0)$. Di conseguenza, in questa condizione, il rapporto di verosimiglianza assume valori molto piccoli, anche prossimi allo zero.

Mediante la statistica G si sottopone a verifica il seguente sistema di ipotesi:

$$H_0: \beta_0 = \beta_1 = \dots = \beta_p = 0$$

$$H_1: \text{almeno un parametro } \beta_j \neq 0, j = 1, \dots, p$$

L'ipotesi nulla H_0 implica che date p variabili indipendenti, nessuna apporti un contributo significativo al modello; mentre per l'ipotesi H_1 almeno una fra le p variabili indipendenti risulta significativa per il modello.

L'ipotesi nulla può essere rifiutata se:

- è soddisfatta la disequazione: $G > \chi_{\frac{\alpha}{2}}^2$; dove il valore $\chi_{\frac{\alpha}{2}}^2$ è tabellato in funzione di α , livello di significatività scelto per il test, e del numero di gradi libertà;
- il valore di G ha un p -value inferiore al livello di significatività α scelto.

Il rifiuto dell'ipotesi nulla implica che la previsione del modello può essere migliorata dalla presenza delle variabili.

5.3.2 Lo Pseudo- R^2

La bontà di adattamento può essere stimata attraverso il coefficiente *Pseudo- R^2* , definito utilizzando il rapporto di verosimiglianza:

$$Pseudo - R_g^2 = 1 - 2 \log \left[\frac{L(0)}{L(\beta)} \right]^{\frac{2}{n}} \quad (14)$$

Dove:

- $Pseudo - R_g^2$ è il coefficiente di determinazione per modelli non lineari;
- $L(0)$ è la massima verosimiglianza in corrispondenza del modello con la sola intercetta;
- $L(\beta)$ è la funzione di massima verosimiglianza in corrispondenza del modello completo;
- n è il numero totale di osservazioni.

Il coefficiente assume valori compresi tra 0 e 1: il limite inferiore prevede che il modello considerato non aggiunga informazioni al modello con la sola intercetta, mentre valori prossimi all'unità implicano un ottimale adattamento del modello ai dati.

Si può far riferimento al coefficiente scalato:

$$\overline{R_g^2} = \frac{R_g^2}{R_{g,\max}^2} \quad (15)$$

Dove il coefficiente di determinazione è scalato rispetto al valore massimo $R_{g,\max}^2$, che è definito come:

$$R_{g,\max}^2 = 1 - [L(0)]^{\frac{2}{n}} \quad (16)$$

Tale metodo non tiene conto del numero di gradi di libertà del modello e non è sottoposto a test di significatività, conseguentemente può essere visto come una misura approssimata.

5.3.3 Test della bontà di adattamento: Devianza, χ^2 di Pearson e Test di Hosmer-Lemeshow

Il programma *Minitab* consente di valutare la bontà di adattamento del modello di regressione logistica tramite altri tre parametri.

Si definisce *devianza* una statistica avente una distribuzione χ^2 con $n-p$ gradi di libertà, dove n è il numero di osservazioni e p è il numero di parametri del modello. Affinché il modello fornisca un'adeguata stima dei dati, il valore di devianza deve essere limitato, al contrario valori grandi di devianza implicano che il modello non sia adeguato. Un modo efficace per valutare la bontà di adattamento è rapportare quest'ultima al numero di gradi di libertà: se tale rapporto supera di molto l'unità, il modello non è adeguato per stimare i dati.

La bontà di adattamento può anche essere valutata con una statistica χ^2 di *Pearson*, che compara le probabilità di successo e fallimento osservate e quelle ottenute di ogni gruppo di osservazioni. Affinché il modello sia adatto a descrivere i dati, devono valere le stesse considerazioni fatte per la statistica devianza, infatti se dividendo la statistica per il numero di gradi libertà $n-p$ si ottiene un valore molto più grande dell'unità il modello non è adeguato.

Il test *Hosmer-Lemeshow (HL)* prevede di classificare in g gruppi i dati in funzione della probabilità di successo stimata; quando $g=10$ i gruppi sono chiamati decili di rischio. Se il campione di dati è sufficientemente grande, la statistica segue una distribuzione del χ^2 con $g-p$ gradi di libertà. Grandi valori di *HL* implicano che il modello non si adatta bene ai dati.

5.4 SIGNIFICATIVITÀ DEI PARAMETRI DEL MODELLO

La valutazione del contributo specifico che ogni regressore apporta sulla stima della variabile dicotomica è interessante al fine di determinare le informazioni che le prime introducono nei riguardi del modello. La valutazione del contributo di ciascuna variabile avviene considerando i coefficienti di regressione β_j e facendo ricorso al test *Z* o alla *statistica di Wald*.

Il test *Z* è definito dalla relazione:

$$Z = \frac{b_j}{s(b_j)} \quad (17)$$

Con:

- b_j stima del parametro β_j ;
- $s(b_j)$ errore standard.

Supposto che valgano le seguenti ipotesi:

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0$$

Z si distribuisce come una distribuzione normale standardizzata.

Se l'esito del test prevede che:

$$|Z| > Z_{\frac{\alpha}{2}}$$

Allora si respinge l'ipotesi nulla H_0 ed è possibile affermare che il parametro è significativamente diverso da 0, di conseguenza la variabile esplicativa corrispondente influisce sulla variabile risposta. La quantità $Z_{\frac{\alpha}{2}}$ rappresenta il valore prelevato dalla tabella della distribuzione normale standard, in base al livello di significatività α scelto.

La *statistica di Wald* coincide con il quadrato di Z e si esprime come:

$$W^2 = \left(\frac{b_j}{s(b_j)} \right)^2 \quad (18)$$

Tale statistica segue una distribuzione χ^2 con 1 grado di libertà.

5.5 ESEMPIO DIMOSTRATIVO

Per chiarire quanto complessivamente scritto a riguardo della regressione logistica, si riporta un esempio tratto dalla letteratura.

Noto il numero di verdetti di pena di morte, si vuole determinare la dipendenza tra la probabilità di condanna e la razza dell'accusato e/o della razza della vittima.

La variabile dicotomica Y assume valore:

- $Y=1$ se è stata assegnata la condanna a morte;
- $Y=0$ se non è stata assegnata la condanna a morte.

Le due variabili indipendenti sono di tipo qualitativo e sono identificate attraverso valori numerici:

- 1 se la razza dell'imputato o della vittima è bianca;

- 0 se la razza dell'imputato o della vittima è nera.

Infine, in tabella è riportata la percentuale di condannati per le quattro combinazioni.

Tabella 5-1 Verdicti di pena di morte secondo la razza dell'imputato e della vittima nei casi di omicidi plurimi in Florida

Razza imputato	Razza vittima	Pena di morte		% Si
		Si	No	
Bianca	Bianca	53	414	11.3
	Nera	0	16	0.0
Nera	Bianca	11	37	22.9
	Nera	4	139	2.8

Analizzando preliminarmente i dati, si osserva che:

- 11 imputati bianchi su 100 vengono condannati a morte nel caso di vittima bianca; 0 imputati bianchi vengono condannati a morte nel caso di vittima nera;
- ad un imputato nero è stata assegnata la condanna a morte nel 22.9% dei casi di vittima bianca e nel 2.8% dei casi di vittima nera.

La costruzione del modello di regressione logistica, prevede che siano definite le variabili:

- Y , pena di morte (si=1; no=0);
- x_1 , razza dell'imputato (si=1; no=0);
- x_2 , razza della vittima (si=1; no=0).

La legge che descrive il modello è:

$$\text{logit}[P(Y = 1)] = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Dove β_0 è la costante del modello, β_1 e β_2 sono l'effetto della razza dell'imputato e della razza della vittima.

Gli *Odds Ratio* per le due variabili indipendenti sono calcolati come:

- $OR_1 = e^{\beta_1}$ che corrisponde al rapporto di odds tra Y e la razza dell'imputato;
- $OR_2 = e^{\beta_2}$ che corrisponde al rapporto di odds tra Y e la razza della vittima.

Per determinare le stime di massima verosimiglianza per la regressione logistica si è utilizzato il software *Gretl* e l'output del programma fornisce: i valori dei parametri, gli errori sulle stime, la significatività di ogni variabile e i valori dei test di bontà di adattamento del modello.

```
Modello 1: Logit multinomiale, usando le osservazioni 1-674
Variabile dipendente: penadimorte
Errori standard basati sull'Hessiana
```

	coefficiente	errore std.	z	p-value	

penadimorte = 1					
const	-3,59610	0,506914	-7,094	1,30e-012	***
razzaimputato	-0,867797	0,367074	-2,364	0,0181	**
razzavittima	2,40444	0,600616	4,003	6,25e-05	***
Media var. dipendente	0,100890	SQM var. dipendente	0,301407		
Log-verosimiglianza	-209,4783	Criterio di Akaike	424,9565		
Criterio di Schwarz	438,4962	Hannan-Quinn	430,1995		
Note: SQM = scarto quadratico medio; E.S. = errore standard					
Numero dei casi 'previsti correttamente' = 606 (89,9%)					
Test del rapporto di verosimiglianza: Chi-quadro(2) = 21,8861 [0,0000]					

Figura 5-2 Output del software *Gretl* per i dati sulla pena di morte

Per i coefficienti di regressione β_j ottenuti, il modello assume la forma:

$$\text{logit}[P(Y = 1)] = -3.5961 - 0.867797x_1 + 2.4044x_2$$

In merito ai segni assunti dai coefficienti β_j è possibile affermare che:

- il segno negativo di β_1 indica che la razza bianca ha una probabilità inferiore di condanna rispetto alla razza nera;
- il segno positivo di β_2 indica che nel caso di vittima bianca, la probabilità di condanna aumenta rispetto al caso di vittima nera.

La definizione della probabilità stimata avviene applicando l'equazione (3) alle osservazioni:

$$p(Y = 1) = \frac{\exp(-3.5961 - 0.867797x_1 + 2.4044x_2)}{1 + \exp(-3.5961 - 0.867797x_1 + 2.4044x_2)}$$

Il valore dell'*Odds Ratio* tra la variabile dipendente e la variabile "razza dell'imputato" è pari a :

$$OR_1 = e^{-0.867797} = 0.42$$

Un imputato bianco ha 0.42 volte l'*Odds*, quindi circa la metà, rispetto ad un imputato nero di essere condannato a morte, cioè la tendenza alla condanna nel primo caso è circa la metà rispetto al secondo.

Il valore dell'*Odds Ratio* tra la variabile dipendente e la variabile “razza della vittima” è pari a:

$$OR_2 = e^{2.4044} = 11.1$$

Ciò significa che l'*Odds* di essere condannato a morte in caso di vittima bianca è 11.1 volte quello nel caso di vittima nera. Quindi, la “propensione alla condanna” nel caso di vittima bianca è 11 volte quella del caso di vittima nera.

La significatività del singolo parametro è esaminata attraverso la *statistica di Wald*. Considerata la variabile indipendente x_1 , per l'ipotesi nulla $H_0 (\beta_1=0)$, il *Test Z* è ottenuto come:

$$Z = \frac{b_j}{s(b_j)} = -\frac{0.868}{0.367} = -2.36$$

Dove il valore $s(b_j)$ è presente nella colonna ‘*errore std.*’ della *Figura 5.2* e il valore Z nella colonna ‘*z*’. Scegliendo un livello di significatività $\alpha=0.05$, si ottiene:

$$|Z| = 2.36 > Z_{\frac{\alpha}{2}} = 1.96$$

La *statistica di Wald* sarà:

$$W^2 = \left[\frac{b_j}{s(b_j)} \right]^2 = 5.59$$

Ed il valore del *p-value* è pari a $0.018 < 0.05$ (livello di significatività α).

Eseguendo gli stessi calcoli per la variabile indipendente x_2 si ottiene:

$$Z = \frac{b_j}{s(b_j)} = \frac{2.404}{0.601} = 4.003$$

$$|Z| = 4.003 > Z_{\frac{\alpha}{2}} = 1.96$$

$$W^2 = \left[\frac{b_j}{s(b_j)} \right]^2 = 16.03$$

Ed il *p-value* è pari a $0.000006 < 0.05$.

Il software restituisce, quindi, il valore di *Z* e del *p-value* ed una volta imposto il livello di significatività, è sufficiente confrontare i valori per stabilire se un effetto è rilevante.

La significatività dei parametri è espressa anche mediante un indicatore visivo contenuto nell'ultima colonna; questo è rappresentato da asterischi presenti in numero crescente, da 0 a 3, in funzione della significatività del parametro.

La bontà di adattamento del modello si testa applicando il test *rapporto di verosimiglianza G*. In Figura 5.2 tale valore è riportato con la dicitura:

Test del rapporto di verosimiglianza: Chi-quadro (2) = 21.8861 [0.0000]

Il numero '2' tra parentesi tonde rappresenta il numero di gradi di libertà del modello. Il valore del *p-value* è riportato tra parentesi quadre e risulta minore di 0.05, conseguentemente l'ipotesi nulla può essere rifiutata.

6 APPLICAZIONE DEL MODELLO DI REGRESSIONE LOGISTICA

Nel seguente capitolo si presentano i passi che hanno costituito l'essenza della costruzione del modello di regressione logistica e della sua validazione e applicazione. In definitiva il modello fornirà dei valori attesi della probabilità di rottura per ogni condotta.

Le variabili prese in considerazione, presenti nella tabella *RegressioneTGM*, chiamate anche 'variabili predittive', e che appaiono influenzare il tasso di rottura delle condotte acquadottistiche, possono essere suddivise in tre categorie:

- ❖ Fattori strutturali:
 - anno di posa;
 - diametro;
 - materiale.
- ❖ Fattori operativi:
 - pressione massima di esercizio;
 - variazione di carico.
- ❖ Fattori ambientali:
 - flusso veicolare;
 - velocità media del flusso;
 - passaggio tram.

La variabile dipendente dicotomica Y assume i valori 1 e 0, a cui corrisponde rispettivamente la presenza e l'assenza di rottura.

Il capitolo è così strutturato:

- ❖ nel §6.1 vengono estratti 20 set di dati dalla tabella *RegressioneTGM*, sui quali viene definito il modello. I set sono costruiti in maniera tale che la percentuale di rottura della tabella *ReteTorino* rimanga costante;
- ❖ nel §6.2 per meglio interpretare il modello si applica il metodo *backward*, riducendo la numerosità delle variabili. Ogni set di dati è stato suddiviso in due parti in un rapporto 80:20; verranno stimati i parametri β delle variabili che effettivamente incidono sulla Y . Nel restante 20% verrà applicata la relazione matematica, che lega i predittori significativi alla variabile dipendente;
- ❖ poiché il valore medio di rottura predetto nelle 20 estrazioni al 20% è sempre maggiore nelle condotte con flag pari a 1, indice del buon funzionamento del modello, nel §6.3 si applica il modello all'intera tabella *RegressioneTGM*.

6.1 ESTRAZIONE DI 20 SET DI DATI E STIMA DEL MODELLO

La tabella *RegressioneTGM* costituisce, quindi, il punto di partenza per la costruzione del modello di regressione logistica e presenta:

- 2548 condotte;
- 2499 condotte non rotte (*flag=0*);
- 49 condotte rotte (*flag=1*);
- 63 rotture;
- percentuale di rottura pari al 2.47%.

Per la definizione del modello si è fatto in modo che la percentuale di rottura rimanesse quella della tabella *ReteTorino*, contenente tutte le condotte della rete, poiché il valore 2.47% è frutto di operazioni di filtraggio dei dati e non rappresenta la vera percentuale di rottura.

La tabella *ReteTorino* è costituita da:

- 25718 condotte;
- 22848 condotte non rotte (*flag=0*);
- 2870 condotte rotte (*flag=1*);
- 3881 rotture;
- percentuale di rottura pari al 15%.

Si riportano i passi eseguiti per poter mantenere una percentuale di rottura pari al 15%:

- sono stati estratti dalla tabella *RegressioneTGM* i 63 guasti (cioè le 49 condotte con *flag=1*);
- sono stati estratti dalla tabella *RegressioneTGM* in maniera casuale 20 campioni da $357 \left(63 * \frac{[(100-15)\%]}{15\%} \right)$ tratti non rotti;
- sono state ottenute 20 estrazioni formate dalle 63 rotture e uno dei 20 campioni da 357 tratti non rotti.

Ciascuno dei 20 set di dati, denominati '*Estrazione_TGMn*', con *n* numero dell'estrazione, presenta:

- 357 tratti con *flag=0*;
- 63 tratti con *flag=1*;
- percentuale di rottura pari al 15%.

Si è scelto di analizzare 20 campioni perché le 357 condotte con flag nullo sono state estratte in maniera casuale e conseguentemente un numero esiguo di campioni da testare sarebbe stato riduttivo. Su ognuna delle venticinque estrazioni è stato tarato il modello di regressione logistica utilizzando il software *Gretl*.

Le variabili del modello sono:

- variabile risposta dipendente Y : flag (1=rottura; 0=non rottura);
- variabile indipendente x_1 : numero materiale;
- variabile indipendente x_2 : diametro;
- variabile indipendente x_3 : anno di posa;
- variabile indipendente x_4 : pressione massima di esercizio;
- variabile indipendente x_5 : velocità media;
- variabile indipendente x_6 : flusso veicolare;
- variabile indipendente x_7 : Tram Route.

Come già precisato, la variabile ‘*variazione di carico*’ è stata esclusa dall’analisi, in quanto correlata con ‘*pressione massima*’, inoltre, poiché la variabile ‘*materiale*’ è di tipo qualitativo, per la regressione è stato assegnato un numero crescente ad ogni tipologia di materiale presente in rete, in funzione del valore del tasso di rottura ottenuto nella tabella *ReteTorino* in ordine decrescente.

Definite le variabili, il modello di regressione logistica assume la forma:

$$\text{logit}[P(Y = 1)] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7$$

La probabilità di rottura stimata è calcolata come:

$$\hat{\pi}(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7}}$$

Vengono riportati i dati di output ottenuti dal software, seguiti da brevi osservazioni, per ogni ‘*Estrazione_TGMn*’. A titolo illustrativo, si riporta la totalità dei passaggi per la definizione delle variabili del modello, la stima dei parametri e la valutazione della bontà di adattamento eseguiti sul set di dati *Estrazione_TGM1*.

Estrazione TGM1

L'output generato da *Gretl* consente di ottenere una stima delle probabilità di rottura.

Il modello di regressione stimato, in funzione dei coefficienti ottenuti, assume la forma:

$$\begin{aligned} \text{logit}[P(Y = 1)] \\ = -17.6820 - 1.4064x_1 - 0.0059x_2 + 0.0099x_3 + 0.0212x_4 \\ + 0.0155x_5 - 0.0005x_6 + 1.1920x_7 \end{aligned}$$

La probabilità di rottura stimata è calcolata come:

$$\hat{\pi}(x) = \frac{e^{-17.6820-1.4064x_1-0.0059x_2+0.0099x_3+0.0212x_4+0.0155x_5-0.0005x_6+1.1920x_7}}{1 + e^{-17.6820-1.4064x_1-0.0059x_2+0.0099x_3+0.0212x_4+0.0155x_5-0.0005x_6+1.1920x_7}}$$

Modello 2: Logit multinomiale, usando le osservazioni 1-420				
Variabile dipendente: flag				
Errori standard basati sull'Hessiana				
	coefficiente	errore std.	z	p-value

flag = 1				
const	-17,6820	12,1795	-1,452	0,1466
N_materiale	-1,40644	0,260193	-5,405	6,47e-08 ***
diametro	-0,00595349	0,00249452	-2,387	0,0170 **
anno_posa	0,00998711	0,00655543	1,523	0,1276
pMax	0,0212333	0,0189732	1,119	0,2631
velocita	0,0155397	0,0364968	0,4258	0,6703
flusso	-0,000543492	0,000846189	-0,6423	0,5207
TramRoute	1,19209	0,413762	2,881	0,0040 ***
Media var. dipendente	0,150000	SQM var. dipendente	0,357497	
Log-verosimiglianza	-117,3332	Criterio di Akaike	250,6665	
Criterio di Schwarz	282,9885	Hannan-Quinn	263,4416	
Note: SQM = scarto quadratico medio; E.S. = errore standard				
Numero dei casi 'previsti correttamente' = 378 (90,0%)				
Test del rapporto di verosimiglianza: Chi-quadro(7) = 120,409 [0,0000]				

Figura 6-1 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM1

Il segno algebrico dei coefficienti fornisce precisazioni, fondamentali e importanti per comprendere la funzionalità del modello:

- il segno negativo di β_1 indica che all'aumentare del numero assegnato a ciascun materiale, la probabilità di rottura diminuisce;
- il segno negativo di β_2 indica che all'aumentare del diametro della condotta, la probabilità di rottura diminuisce;

- il segno positivo di β_3 indica che all'aumentare dell'anno di posa (quindi quanto più una condotta è recente), la probabilità di rottura aumenta. Condotte più giovani hanno maggiore probabilità di rottura;
- il segno positivo di β_4 indica che all'aumentare della pressione massima di esercizio della condotta, la probabilità di rottura aumenta;
- il segno positivo di β_5 indica che all'aumentare della velocità media dei carichi dinamici sovrastanti la condotta, la probabilità di rottura aumenta;
- il segno negativo di β_6 indica che all'aumentare del flusso veicolare, la probabilità di rottura diminuisce;
- il segno positivo di β_7 indica che la presenza del tram fa sì che la probabilità di rottura aumenti.

Per valutare la significatività di ogni variabile predittiva è sufficiente verificare che i valori contenuti nella colonna 'z', relativi al *Test Z*, definito un livello di significatività α pari a 0.05, soddisfino la relazione:

$$|Z| > Z_{\frac{\alpha}{2}} = 1.96$$

Quest'ultima è soddisfatta per tre variabili su sette, come confermano i valori del *p-value* e l'indicatore visivo. Si ricorda, infatti, che *Gretl* fornisce nell'ultima colonna un numero di asterischi che varia tra 0 e 3, in funzione della significatività di ogni singolo parametro; si riporta l'interpretazione dell'indicatore visivo in Tabella 6-1.

Tabella 6-1 Indicatore visivo associato alla significatività osservata delle variabili

***	<i>p-value</i> < 0,01
**	<i>p-value</i> < 0,05
*	<i>p-value</i> < 0,10

Inoltre il software mostra il valore della statistica *Chi-quadro* alla voce '*Test del rapporto verosimiglianze*'. Il valore restituito dal modello, con 7 gradi di libertà, riportati fra parentesi tonde, soddisfa la relazione:

$$G = 120.409 > \chi^2_{\frac{\alpha}{2}} = 12.38$$

Il valore del *p-value*, riportato fra parentesi quadre, è minore di 0.05. Conseguentemente l'ipotesi nulla può essere rifiutata ed è possibile affermare che il modello completo aggiunge informazioni rispetto al modello con la sola intercetta e che la previsione viene implementata se vengono prese in considerazione le variabili:

- materiale;
- diametro;
- tram route.

I passaggi descritti, eseguiti sul set di dati *Estrazione_TGM1*, sono stati applicati alle restanti 19 estrazioni; se ne riportano i risultati.

Estrazione TGM2

L'output fornito da *Gretl* permette di definire il modello di regressione nella seguente forma:

$$\begin{aligned} \text{logit}[P(Y = 1)] \\ = -27.1752 - 1.44533x_1 - 0.0047x_2 + 0.0153x_3 + 0.0023x_4 \\ + 0.0086x_5 - 0.0005x_6 + 1.1395x_7 \end{aligned}$$

E la probabilità di rottura diviene:

$$\hat{\pi}(x) = \frac{e^{-27.1752-1.44533x_1-0.0047x_2+0.0153x_3+0.0023x_4+0.0086x_5-0.0005x_6+1.1395x_7}}{1 + e^{-27.1752-1.44533x_1-0.0047x_2+0.0153x_3+0.0023x_4+0.0086x_5-0.0005x_6+1.1395x_7}}$$

```
Modello 1: Logit multinomiale, usando le osservazioni 1-420
Variabile dipendente: flag
Errori standard basati sull'Hessiana
```

	coefficiente	errore std.	z	p-value	

flag = 1					
const	-27,1752	11,6211	-2,338	0,0194	**
N_materiale	-1,44533	0,255103	-5,666	1,46e-08	***
diámetro	-0,00472162	0,00217330	-2,173	0,0298	**
anno_posa	0,0153038	0,00623091	2,456	0,0140	**
pMax	0,00232753	0,0120174	0,1937	0,8464	
velocità	0,00865959	0,0306500	0,2825	0,7775	
flusso	-0,000520641	0,000748172	-0,6959	0,4865	
TramRoute	1,13958	0,376779	3,025	0,0025	***
Media var. dipendente	0,150000	SQM var. dipendente	0,357497		
Log-verosimiglianza	-126,5208	Criterio di Akaike	269,0416		
Criterio di Schwarz	301,3637	Hannan-Quinn	281,8168		
Note: SQM = scarto quadratico medio; E.S. = errore standard					
Numero dei casi 'previsti correttamente' = 375 (89,3%)					
Test del rapporto di verosimiglianza: Chi-quadro(7) = 102,034 [0,0000]					

Figura 6-2 Stime dei parametri del modello di regressione logistica applicato al set di dati *Estrazione_TGM2*

Si traggono le seguenti considerazioni:

- il segno negativo di β_1 indica che all'aumentare del numero assegnato a ciascun materiale, la probabilità di rottura diminuisce;
- il segno negativo di β_2 indica che all'aumentare del diametro della condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_3 indica che all'aumentare dell'anno di posa (quindi quanto più una condotta è recente), la probabilità di rottura aumenta. Condotte più giovani hanno maggiore probabilità di rottura;
- il segno positivo di β_4 indica che all'aumentare della pressione massima di esercizio della condotta, la probabilità di rottura aumenta;
- il segno positivo di β_5 indica che all'aumentare della velocità media dei carichi dinamici sovrastanti la condotta, la probabilità di rottura aumenta;
- il segno negativo di β_6 indica che all'aumentare del flusso veicolare, la probabilità di rottura diminuisce;
- il segno positivo di β_7 indica che la presenza del tram fa sì che la probabilità di rottura aumenti.

I valori del *Test di Z* relativi ad ogni variabile, fissato un livello di significatività α pari a 0.05, mostrano che variabili quali: pressione massima, velocità media e flusso non risultano significative.

Inoltre il software mostra il valore della statistica *Chi-quadro* alla voce '*Test del rapporto verosimiglianze*'. Il valore restituito dal modello, con 7 gradi di libertà, riportati fra parentesi tonde, soddisfa la relazione:

$$G = 102.034 > \chi^2_{\frac{\alpha}{2}} = 12.38$$

Conseguentemente l'ipotesi nulla può essere rifiutata ed è possibile affermare che il modello completo aggiunge informazioni rispetto al modello con la sola intercetta e che la previsione viene implementata se vengono prese in considerazione le variabili:

- materiale;
- diametro;
- anno di posa;
- tram route.

Estrazione TGM3

Si riporta l'output fornito da *Gretl*. Il modello di regressione stimato, in funzione dei coefficienti ottenuti, assume la forma:

$$\begin{aligned} \text{logit}[P(Y = 1)] \\ = -2.5954 - 1.0797x_1 - 0.0068x_2 + 0.0022x_3 + 0.0088x_4 - 0.0061x_5 \\ + 0.0005x_6 + 0.9713x_7 \end{aligned}$$

La probabilità di rottura stimata è calcolata come:

$$\hat{\pi}(x) = \frac{e^{-2.5954-1.0797x_1-0.0068x_2+0.0022x_3+0.0088x_4-0.0061x_5+0.0005x_6+0.9713x_7}}{1 + e^{-2.5954-1.0797x_1-0.0068x_2+0.0022x_3+0.0088x_4-0.0061x_5+0.0005x_6+0.9713x_7}}$$

Modello 1: Logit multinomiale, usando le osservazioni 1-420				
Variabile dipendente: flag				
Errori standard basati sull'Hessiana				
	coefficiente	errore std.	z	p-value

flag = 1				
const	-2,59545	10,7762	-0,2409	0,8097
N_materiale	-1,07977	0,232935	-4,635	3,56e-06 ***
diametro	-0,00688809	0,00274098	-2,513	0,0120 **
anno_posa	0,00221101	0,00575317	0,3843	0,7007
pMax	0,00882303	0,0113056	0,7804	0,4351
velocita	-0,00612279	0,0370807	-0,1651	0,8688
flusso	0,000500824	0,000781191	0,6411	0,5215
TramRoute	0,971335	0,376910	2,577	0,0100 ***
Media var. dipendente	0,150000	SQM var. dipendente	0,357497	
Log-verosimiglianza	-121,7236	Criterio di Akaike	259,4472	
Criterio di Schwarz	291,7692	Hannan-Quinn	272,2223	
Note: SQM = scarto quadratico medio; E.S. = errore standard				
Numero dei casi 'previsti correttamente' = 375 (89,3%)				
Test del rapporto di verosimiglianza: Chi-quadro(7) = 111,628 [0,0000]				

Figura 6-3 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM3

Il segno algebrico dei coefficienti fornisce precisazioni, fondamentali e importanti per comprendere la funzionalità del modello:

- il segno negativo di β_1 indica che all'aumentare del numero assegnato a ciascun materiale, la probabilità di rottura diminuisce;
- il segno negativo di β_2 indica che all'aumentare del diametro della condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_3 indica che all'aumentare dell'anno di posa (quindi quanto più una condotta è recente), la probabilità di rottura aumenta. Condotte più giovani hanno maggiore probabilità di rottura;
- il segno positivo di β_4 indica che all'aumentare della pressione massima di esercizio della condotta, la probabilità di rottura aumenta;

- il segno negativo di β_5 indica che all'aumentare della velocità media dei carichi dinamici sovrastanti la condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_6 indica che all'aumentare del flusso veicolare, la probabilità di rottura aumenta;
- il segno positivo di β_7 indica che la presenza del tram fa sì che la probabilità di rottura aumenti.

I valori del *Test di Z* relativi ad ogni variabile, fissato un livello di significatività α pari a 0.05, mostrano che variabili quali: anno di posa, pressione massima, velocità media e flusso non risultano significative.

Inoltre il software mostra il valore della statistica *Chi-quadro* alla voce '*Test del rapporto verosimiglianze*'. Il valore restituito dal modello, con 7 gradi di libertà, riportati fra parentesi tonde, soddisfa la relazione:

$$G = 111.628 > \chi^2_{\frac{\alpha}{2}} = 12.38$$

Conseguentemente l'ipotesi nulla può essere rifiutata ed è possibile affermare che il modello completo aggiunge informazioni rispetto al modello con la sola intercetta e che la previsione viene implementata se vengono prese in considerazione le variabili:

- materiale;
- diametro;
- tram route.

Estrazione TGM4

L'output fornito da *Gretl* permette di definire il modello di regressione nella seguente forma:

$$\begin{aligned} \text{logit}[P(Y = 1)] \\ = 1.8032 - 1.1256x_1 - 0.0055x_2 + 0.0003x_3 + 0.0030x_4 - 0.0157x_5 \\ + 0.0003x_6 + 0.7961x_7 \end{aligned}$$

E la probabilità di rottura diviene:

$$\hat{\pi}(x) = \frac{e^{1.8032 - 1.1256x_1 - 0.0055x_2 + 0.0003x_3 + 0.0030x_4 - 0.0157x_5 + 0.0003x_6 + 0.7961x_7}}{1 + e^{1.8032 - 1.1256x_1 - 0.0055x_2 + 0.0003x_3 + 0.0030x_4 - 0.0157x_5 + 0.0003x_6 + 0.7961x_7}}$$

```

Modello 1: Logit multinomiale, usando le osservazioni 1-420
Variabile dipendente: flag
Errori standard basati sull'Hessiana

      coefficiente      errore std.      z      p-value
-----
flag = 1
const      1,80328      10,8212      0,1666      0,8677
N_materiale -1,12563      0,235434     -4,781      1,74e-06 ***
diametro   -0,00556449      0,00275929    -2,017      0,0437 **
anno_posa   0,000321312      0,00574219     0,05596     0,9554
pMax        0,00307192      0,00812137     0,3783      0,7052
velocitaA  -0,0157182      0,0349312     -0,4500      0,6527
flusso      0,000387073      0,000811995     0,4767      0,6336
TramRoute   0,796174      0,381788      2,085      0,0370 **

Media var. dipendente  0,150000      SQM var. dipendente  0,357497
Log-verosimiglianza    -115,4153      Criterio di Akaike    246,8306
Criterio di Schwarz     279,1527      Hannan-Quinn          259,6058
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 378 (90,0%)
Test del rapporto di verosimiglianza: Chi-quadro(7) = 124,245 [0,0000]

```

Figura 6-4 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM4

I coefficienti ottenuti consentono di affermare che:

- il segno negativo di β_1 indica che all'aumentare del numero assegnato a ciascun materiale, la probabilità di rottura diminuisce;
- il segno negativo di β_2 indica che all'aumentare del diametro della condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_3 indica che all'aumentare dell'anno di posa (quindi quanto più una condotta è recente), la probabilità di rottura aumenta. Condotte più giovani hanno maggiore probabilità di rottura;
- il segno positivo di β_4 indica che all'aumentare della pressione massima di esercizio della condotta, la probabilità di rottura aumenta;
- il segno negativo di β_5 indica che all'aumentare della velocità media dei carichi dinamici sovrastanti la condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_6 indica che all'aumentare del flusso veicolare, la probabilità di rottura aumenta;
- il segno positivo di β_7 indica che la presenza del tram fa sì che la probabilità di rottura aumenti.

I valori del *Test di Z* relativi ad ogni variabile, fissato un livello di significatività α pari a 0.05, mostrano che variabili quali: anno di posa, pressione massima, velocità media e flusso non risultano significative.

Inoltre il software mostra il valore della statistica *Chi-quadro* alla voce ‘*Test del rapporto verosimiglianze*’. Il valore restituito dal modello, con 7 gradi di libertà, riportati fra parentesi tonde, soddisfa la relazione:

$$G = 124.245 > \chi^2_{\frac{\alpha}{2}} = 12.38$$

Conseguentemente l’ipotesi nulla può essere rifiutata ed è possibile affermare che il modello completo aggiunge informazioni rispetto al modello con la sola intercetta e che la previsione viene implementata se vengono prese in considerazione le variabili:

- materiale;
- diametro;
- tram route.

Estrazione TGM5

Si riporta l’output fornito da *Gretl*. Il modello di regressione stimato, in funzione dei coefficienti ottenuti, assume la forma:

$$\begin{aligned} \text{logit}[P(Y = 1)] \\ = 4.3359 - 1.0142x_1 - 0.0045x_2 - 0.0014x_3 + 0.0046x_4 + 0.0017x_5 \\ + 0.0002x_6 + 0.6556x_7 \end{aligned}$$

La probabilità di rottura stimata è calcolata come:

$$\hat{\pi}(x) = \frac{e^{4.3359-1.0142x_1-0.0045x_2-0.0014x_3+0.0046x_4+0.0017x_5+0.0002x_6+0.6556x_7}}{1 + e^{4.3359-1.0142x_1-0.0045x_2-0.0014x_3+0.0046x_4+0.0017x_5+0.0002x_6+0.6556x_7}}$$

Il segno algebrico dei coefficienti fornisce precisazioni, fondamentali e importanti per comprendere la funzionalità del modello:

- il segno negativo di β_1 indica che all’aumentare del numero assegnato a ciascun materiale, la probabilità di rottura diminuisce;
- il segno negativo di β_2 indica che all’aumentare del diametro della condotta, la probabilità di rottura diminuisce;
- il segno negativo di β_3 indica che all’aumentare dell’anno di posa (quindi quanto più una condotta è recente), la probabilità di rottura diminuisce. Condotte più giovani hanno minore probabilità di rottura;
- il segno positivo di β_4 indica che all’aumentare della pressione massima di esercizio della condotta, la probabilità di rottura aumenta;

- il segno positivo di β_5 indica che all'aumentare della velocità media dei carichi dinamici sovrastanti la condotta, la probabilità di rottura aumenta;
- il segno positivo di β_6 indica che all'aumentare del flusso veicolare, la probabilità di rottura aumenta;
- il segno positivo di β_7 indica che la presenza del tram fa sì che la probabilità di rottura aumenti.

Modello 1: Logit multinomiale, usando le osservazioni 1-420
 Variabile dipendente: flag
 Errori standard basati sull'Hessiana

	coefficiente	errore std.	z	p-value	

flag = 1					
const	4,33592	11,2753	0,3846	0,7006	
N_materiale	-1,01420	0,240647	-4,214	2,50e-05	***
diametro	-0,00458369	0,00250317	-1,831	0,0671	*
anno_posa	-0,00142123	0,00602412	-0,2359	0,8135	
pMax	0,00462774	0,00685072	0,6755	0,4994	
velocità	0,00172808	0,0343045	0,05037	0,9598	
flusso	0,000299193	0,000769903	0,3886	0,6976	
TramRoute	0,655693	0,380338	1,724	0,0847	*
Media var. dipendente	0,150000	SQM var. dipendente		0,357497	
Log-verosimiglianza	-121,6756	Criterio di Akaike		259,3511	
Criterio di Schwarz	291,6732	Hannan-Quinn		272,1263	
Note: SQM = scarto quadratico medio; E.S. = errore standard					
Numero dei casi 'previsti correttamente' = 372 (88,6%)					
Test del rapporto di verosimiglianza: Chi-quadro(7) = 111,725 [0,0000]					

Figura 6-5 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM5

I valori del *Test di Z* relativi ad ogni variabile, fissato un livello di significatività α pari a 0.05, mostrano che tutte le variabili, eccezion fatta per il 'materiale', non risultano significative.

Inoltre il software mostra il valore della statistica *Chi-quadro* alla voce '*Test del rapporto verosimiglianze*'. Il valore restituito dal modello, con 7 gradi di libertà, riportati fra parentesi tonde, soddisfa la relazione:

$$G = 111.725 > \chi^2_{\alpha/2} = 12.38$$

Conseguentemente l'ipotesi nulla può essere rifiutata ed è possibile affermare che il modello completo aggiunge informazioni rispetto al modello con la sola intercetta e che la previsione viene implementata se viene presa in considerazione la variabile:

- materiale.

Estrazione TGM6

L'output fornito da *Gretl* permette di definire il modello di regressione nella seguente forma:

$$\begin{aligned} \text{logit}[P(Y = 1)] \\ = -12.2218 - 1.2204x_1 - 0.0064x_2 + 0.0074x_3 - 0.0032x_4 \\ + 0.0312x_5 - 0.0005x_6 + 0.9963x_7 \end{aligned}$$

E la probabilità di rottura diviene:

$$\hat{\pi}(x) = \frac{e^{-12.2218-1.2204x_1-0.0064x_2+0.0074x_3-0.0032x_4+0.0312x_5-0.0005x_6+0.9963x_7}}{1 + e^{-12.2218-1.2204x_1-0.0064x_2+0.0074x_3-0.0032x_4+0.0312x_5-0.0005x_6+0.9963x_7}}$$

Si traggono le seguenti conclusioni:

- il segno negativo di β_1 indica che all'aumentare del numero assegnato a ciascun materiale, la probabilità di rottura diminuisce;
- il segno negativo di β_2 indica che all'aumentare del diametro della condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_3 indica che all'aumentare dell'anno di posa (quindi quanto più una condotta è recente), la probabilità di rottura aumenta. Condotte più giovani hanno maggiore probabilità di rottura;
- il segno negativo di β_4 indica che all'aumentare della pressione massima di esercizio della condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_5 indica che all'aumentare della velocità media dei carichi dinamici sovrastanti la condotta, la probabilità di rottura aumenta;
- il segno negativo di β_6 indica che all'aumentare del flusso veicolare, la probabilità di rottura diminuisce;
- il segno positivo di β_7 indica che la presenza del tram fa sì che la probabilità di rottura aumenti.

```

Modello 1: Logit multinomiale, usando le osservazioni 1-420
Variabile dipendente: flag
Errori standard basati sull'Hessiana

      coefficiente      errore std.      z      p-value
-----
flag = 1
const      -12,2218      11,6539      -1,049      0,2943
N_materiale -1,22044      0,251300      -4,857      1,19e-06 ***
diametro    -0,00648793    0,00251062     -2,584      0,0098 ***
anno_posa    0,00744740      0,00623559      1,194      0,2323
pMax         -0,00321239      0,00808024     -0,3976     0,6910
velocita     0,0312073      0,0329693       0,9466     0,3439
flusso       -0,000563282      0,000738055    -0,7632     0,4453
TramRoute    0,996343      0,377801       2,637      0,0084 ***

Media var. dipendente 0,150000      SQM var. dipendente 0,357497
Log-verosimiglianza -123,0659      Criterio di Akaike 262,1318
Criterio di Schwarz 294,4538      Hannan-Quinn 274,9069
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 375 (89,3%)
Test del rapporto di verosimiglianza: Chi-quadro(7) = 108,944 [0,0000]

```

Figura 6-6 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM6

I valori del *Test di Z* relativi ad ogni variabile, fissato un livello di significatività α pari a 0.05, mostrano che variabili quali: anno di posa, pressione massima, velocità media e flusso non risultano significative.

Inoltre il software mostra il valore della statistica *Chi-quadro* alla voce '*Test del rapporto verosimiglianze*'. Il valore restituito dal modello, con 7 gradi di libertà, riportati fra parentesi tonde, soddisfa la relazione:

$$G = 108.944 > \chi_{\frac{\alpha}{2}}^2 = 12.38$$

Conseguentemente l'ipotesi nulla può essere rifiutata ed è possibile affermare che il modello completo aggiunge informazioni rispetto al modello con la sola intercetta e che la previsione viene implementata se vengono prese in considerazione le variabili:

- materiale;
- diametro;
- tram route.

Estrazione TGM7

Si riporta l'output fornito da *Gretl*. Il modello di regressione stimato, in funzione dei coefficienti ottenuti, assume la forma:

$$\begin{aligned} \text{logit}[P(Y = 1)] \\ = -5.3046 - 1.1338x_1 - 0.0044x_2 + 0.0037x_3 + 0.0051x_4 - 0.0073x_5 \\ - 0.00007x_6 + 0.6312x_7 \end{aligned}$$

La probabilità di rottura stimata è calcolata come:

$$\hat{\pi}(x) = \frac{e^{-5.3046-1.1338x_1-0.0044x_2+0.0037x_3+0.0051x_4-0.0073x_5-0.00007x_6+0.6312x_7}}{1 + e^{-5.3046-1.1338x_1-0.0044x_2+0.0037x_3+0.0051x_4-0.0073x_5-0.00007x_6+0.6312x_7}}$$

```
Modello 1: Logit multinomiale, usando le osservazioni 1-420
Variabile dipendente: flag
Errori standard basati sull'Hessiana
```

	coefficiente	errore std.	z	p-value

flag = 1				
const	-5,30463	10,7087	-0,4954	0,6203
N_materiale	-1,13381	0,233425	-4,857	1,19e-06 ***
diametro	-0,00443528	0,00228298	-1,943	0,0520 *
anno_posa	0,00378569	0,00571315	0,6626	0,5076
pMax	0,00516238	0,00825661	0,6252	0,5318
velocita	-0,00735420	0,0318151	-0,2312	0,8172
flusso	-7,07587e-05	0,000769084	-0,09200	0,9267
TramRoute	0,631282	0,369327	1,709	0,0874 *
Media var. dipendente	0,150000	SQM var. dipendente		0,357497
Log-verosimiglianza	-126,2081	Criterio di Akaike		268,4162
Criterio di Schwarz	300,7382	Hannan-Quinn		281,1913
Note: SQM = scarto quadratico medio; E.S. = errore standard				
Numero dei casi 'previsti correttamente' = 368 (87,6%)				
Test del rapporto di verosimiglianza: Chi-quadro(7) = 102,659 [0,0000]				

Figura 6-7 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM7

Il segno algebrico dei coefficienti fornisce precisazioni, fondamentali e importanti per comprendere la funzionalità del modello:

- il segno negativo di β_1 indica che all'aumentare del numero assegnato a ciascun materiale, la probabilità di rottura diminuisce;
- il segno negativo di β_2 indica che all'aumentare del diametro della condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_3 indica che all'aumentare dell'anno di posa (quindi quanto più una condotta è recente), la probabilità di rottura aumenta. Condotte più giovani hanno maggiore probabilità di rottura;
- il segno positivo di β_4 indica che all'aumentare della pressione massima di esercizio della condotta, la probabilità di rottura aumenta;

- il segno negativo di β_5 indica che all'aumentare della velocità media dei carichi dinamici sovrastanti la condotta, la probabilità di rottura diminuisce;
- il segno negativo di β_6 indica che all'aumentare del flusso veicolare, la probabilità di rottura diminuisce;
- il segno positivo di β_7 indica che la presenza del tram fa sì che la probabilità di rottura aumenti.

I valori del *Test di Z* relativi ad ogni variabile, fissato un livello di significatività α pari a 0.05, mostrano che tutte le variabili, eccezion fatta per '*materiale*', non risultano significative.

Inoltre il software mostra il valore della statistica *Chi-quadro* alla voce '*Test del rapporto verosimiglianze*'. Il valore restituito dal modello, con 7 gradi di libertà, riportati fra parentesi tonde, soddisfa la relazione:

$$G = 102.659 > \chi^2_{\frac{\alpha}{2}} = 12.38$$

Conseguentemente l'ipotesi nulla può essere rifiutata ed è possibile affermare che il modello completo aggiunge informazioni rispetto al modello con la sola intercetta e che la previsione viene implementata se vengono prese in considerazione le variabili:

- materiale.

Estrazione TGM8

L'output fornito da *Gretl* permette di definire il modello di regressione nella seguente forma:

$$\begin{aligned} \text{logit}[P(Y = 1)] \\ = -12.6923 - 1.2582x_1 - 0.0047x_2 + 0.0076x_3 + 0.0017x_4 \\ + 0.0132x_5 - 0.0006x_6 + 1.0888x_7 \end{aligned}$$

E la probabilità di rottura diviene:

$$\hat{\pi}(x) = \frac{e^{-12.6923-1.2582x_1-0.0047x_2+0.0076x_3+0.0017x_4+0.0132x_5-0.0006x_6+1.0888x_7}}{1 + e^{-12.6923-1.2582x_1-0.0047x_2+0.0076x_3+0.0017x_4+0.0132x_5-0.0006x_6+1.0888x_7}}$$

I coefficienti ottenuti consentono di trarre le seguenti conclusioni:

- il segno negativo di β_1 indica che all'aumentare del numero assegnato a ciascun materiale, la probabilità di rottura diminuisce;

- il segno negativo di β_2 indica che all'aumentare del diametro della condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_3 indica che all'aumentare dell'anno di posa (quindi quanto più una condotta è recente), la probabilità di rottura aumenta. Condotte più giovani hanno maggiore probabilità di rottura;
- il segno positivo di β_4 indica che all'aumentare della pressione massima di esercizio della condotta, la probabilità di rottura aumenta;
- il segno positivo di β_5 indica che all'aumentare della velocità media dei carichi dinamici sovrastanti la condotta, la probabilità di rottura aumenta;
- il segno negativo di β_6 indica che all'aumentare del flusso veicolare, la probabilità di rottura diminuisce;
- il segno positivo di β_7 indica che la presenza del tram fa sì che la probabilità di rottura aumenti.

Modello 1: Logit multinomiale, usando le osservazioni 1-420
 Variabile dipendente: flag
 Errori standard basati sull'Hessiana

	coefficiente	errore std.	z	p-value	
flag = 1					
const	-12,6923	11,1002	-1,143	0,2529	
N_materiale	-1,25827	0,248067	-5,072	3,93e-07	***
diámetro	-0,00479612	0,00259980	-1,845	0,0651	*
anno_posa	0,00769191	0,00592402	1,298	0,1941	
pMax	0,00173420	0,00982298	0,1765	0,8599	
velocità	0,0132576	0,0298601	0,4440	0,6570	
flusso	-0,000693860	0,000822515	-0,8436	0,3989	
TramRoute	1,08886	0,394705	2,759	0,0058	***
Media var. dipendente	0,150000	SQM var. dipendente		0,357497	
Log-verosimiglianza	-123,2061	Criterio di Akaike		262,4123	
Criterio di Schwarz	294,7343	Hannan-Quinn		275,1874	

Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 379 (90,2%)
 Test del rapporto di verosimiglianza: Chi-quadro(7) = 108,663 [0,0000]

Figura 6-8 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM8

I valori del *Test di Z* relativi ad ogni variabile, fissato un livello di significatività α pari a 0.05, mostrano che variabili quali: anno di posa, diametro, pressione massima, velocità media e flusso non risultano significative.

Inoltre il software mostra il valore della statistica *Chi-quadro* alla voce '*Test del rapporto verosimiglianze*'. Il valore restituito dal modello, con 7 gradi di libertà, riportati fra parentesi tonde, soddisfa la relazione:

$$G = 108.663 > \chi^2_{\frac{\alpha}{2}} = 12.38$$

Conseguentemente l'ipotesi nulla può essere rifiutata ed è possibile affermare che il modello completo aggiunge informazioni rispetto al modello con la sola intercetta e che la previsione viene implementata se vengono prese in considerazione le variabili:

- materiale;
- tram route.

Estrazione TGM9

Si riporta l'output fornito da *Gretl*. Il modello di regressione stimato, in funzione dei coefficienti ottenuti, assume la forma:

$$\begin{aligned} \text{logit}[P(Y = 1)] \\ = -8.1957 - 1.4256x_1 - 0.0061x_2 + 0.0056x_3 + 0.0042x_4 - 0.0067x_5 \\ + 0.0008x_6 + 1.0013x_7 \end{aligned}$$

La probabilità di rottura stimata è calcolata come:

$$\hat{\pi}(x) = \frac{e^{-8.1957-1.4256x_1-0.0061x_2+0.0056x_3+0.0042x_4-0.0067x_5+0.0008x_6+1.0013x_7}}{1 + e^{-8.1957-1.4256x_1-0.0061x_2+0.0056x_3+0.0042x_4-0.0067x_5+0.0008x_6+1.0013x_7}}$$

Modello 1: Logit multinomiale, usando le osservazioni 1-420				
Variabile dipendente: flag				
Errori standard basati sull'Hessiana				
	coefficiente	errore std.	z	p-value

flag = 1				
const	-8,19579	12,4108	-0,6604	0,5090
N_materiale	-1,42560	0,272199	-5,237	1,63e-07 ***
diametro	-0,00612891	0,00236540	-2,591	0,0096 ***
anno_posa	0,00567914	0,00664761	0,8543	0,3929
pMax	0,00421766	0,00982862	0,4291	0,6678
velocita	-0,00674798	0,0378682	-0,1782	0,8586
flusso	0,000823056	0,000899877	0,9146	0,3604
TramRoute	1,00134	0,408152	2,453	0,0142 **
Media var. dipendente	0,150000	SQM var. dipendente	0,357497	
Log-verosimiglianza	-109,3369	Criterio di Akaike	234,6738	
Criterio di Schwarz	266,9958	Hannan-Quinn	247,4489	
Note: SQM = scarto quadratico medio; E.S. = errore standard				
Numero dei casi 'previsti correttamente' = 385 (91,7%)				
Test del rapporto di verosimiglianza: Chi-quadro(7) = 136,402 [0,0000]				

Figura 6-9 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM9

Il segno algebrico dei coefficienti fornisce precisazioni, fondamentali e importanti per comprendere la funzionalità del modello:

- il segno negativo di β_1 indica che all'aumentare del numero assegnato a ciascun materiale, la probabilità di rottura diminuisce;
- il segno negativo di β_2 indica che all'aumentare del diametro della condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_3 indica che all'aumentare dell'anno di posa (quindi quanto più una condotta è recente), la probabilità di rottura aumenta. Condotte più giovani hanno maggiore probabilità di rottura;
- il segno positivo di β_4 indica che all'aumentare della pressione massima di esercizio della condotta, la probabilità di rottura aumenta;
- il segno negativo di β_5 indica che all'aumentare della velocità media dei carichi dinamici sovrastanti la condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_6 indica che all'aumentare del flusso veicolare, la probabilità di rottura aumenta;
- il segno positivo di β_7 indica che la presenza del tram fa sì che la probabilità di rottura aumenti.

I valori del *Test di Z* relativi ad ogni variabile, fissato un livello di significatività α pari a 0.05, mostrano che le variabili: anno di posa, pressione massima, velocità media e flusso non risultano significative.

Inoltre il software mostra il valore della statistica *Chi-quadro* alla voce '*Test del rapporto verosimiglianze*'. Il valore restituito dal modello, con 7 gradi di libertà, riportati fra parentesi tonde, soddisfa la relazione:

$$G = 136.402 > \chi^2_{\alpha/2} = 12.38$$

Conseguentemente l'ipotesi nulla può essere rifiutata ed è possibile affermare che il modello completo aggiunge informazioni rispetto al modello con la sola intercetta e che la previsione viene implementata se viene presa in considerazione la variabile:

- materiale;
- diametro;
- tram route.

Estrazione TGM10

Si riporta l'output fornito da *Gretl*. Il modello di regressione stimato, in funzione dei coefficienti ottenuti, assume la forma:

$$\begin{aligned} \text{logit}[P(Y = 1)] \\ = -5.6270 - 1.0722x_1 - 0.0077x_2 + 0.0039x_3 + 0.0015x_4 - 0.0009x_5 \\ + 0.0005x_6 + 0.7323x_7 \end{aligned}$$

La probabilità di rottura stimata è calcolata come:

$$\hat{\pi}(x) = \frac{e^{-5.6270-1.0722x_1-0.0077x_2+0.0039x_3+0.0015x_4-0.0009x_5+0.0005x_6+0.7323x_7}}{1 + e^{-5.6270-1.0722x_1-0.0077x_2+0.0039x_3+0.0015x_4-0.0009x_5+0.0005x_6+0.7323x_7}}$$

Modello 1: Logit multinomiale, usando le osservazioni 1-420
 Variabile dipendente: flag
 Errori standard basati sull'Hessiana

	coefficiente	errore std.	z	p-value

flag = 1				
const	-5,62707	10,6567	-0,5280	0,5975
N_materiale	-1,07226	0,229394	-4,674	2,95e-06 ***
diametro	-0,00774464	0,00299455	-2,586	0,0097 ***
anno_posa	0,00394780	0,00566728	0,6966	0,4861
pMax	0,00155837	0,00846751	0,1840	0,8540
velocita	-0,000906696	0,0358845	-0,02527	0,9798
flusso	0,000584662	0,000768501	0,7608	0,4468
TramRoute	0,732389	0,382826	1,913	0,0557 *
Media var. dipendente	0,150000	SQM var. dipendente	0,357497	
Log-verosimiglianza	-125,1421	Criterio di Akaike	266,2842	
Criterio di Schwarz	298,6062	Hannan-Quinn	279,0593	
Note: SQM = scarto quadratico medio; E.S. = errore standard				
Numero dei casi 'previsti correttamente' = 374 (89,0%)				
Test del rapporto di verosimiglianza: Chi-quadro(7) = 104,791 [0,0000]				

Figura 6-10 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM10

Il segno algebrico dei coefficienti fornisce precisazioni, fondamentali e importanti per comprendere la funzionalità del modello:

- il segno negativo di β_1 indica che all'aumentare del numero assegnato a ciascun materiale, la probabilità di rottura diminuisce;
- il segno negativo di β_2 indica che all'aumentare del diametro della condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_3 indica che all'aumentare dell'anno di posa (quindi quanto più una condotta è recente), la probabilità di rottura aumenta. Condotte più giovani hanno maggiore probabilità di rottura;
- il segno positivo di β_4 indica che all'aumentare della pressione massima di esercizio della condotta, la probabilità di rottura aumenta;

- il segno negativo di β_5 indica che all'aumentare della velocità media dei carichi dinamici sovrastanti la condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_6 indica che all'aumentare del flusso veicolare, la probabilità di rottura aumenta;
- il segno positivo di β_7 indica che la presenza del tram fa sì che la probabilità di rottura aumenti.

I valori del *Test di Z* relativi ad ogni variabile, fissato un livello di significatività α pari a 0.05, mostrano che le variabili: anno di posa, pressione massima, velocità media, flusso e tram non risultano significative.

Inoltre il software mostra il valore della statistica *Chi-quadro* alla voce '*Test del rapporto verosimiglianze*'. Il valore restituito dal modello, con 7 gradi di libertà, riportati fra parentesi tonde, soddisfa la relazione:

$$G = 104.791 > \chi^2_{\frac{\alpha}{2}} = 12.38$$

Conseguentemente l'ipotesi nulla può essere rifiutata ed è possibile affermare che il modello completo aggiunge informazioni rispetto al modello con la sola intercetta e che la previsione viene implementata se viene presa in considerazione la variabile:

- materiale;
- diametro.

Estrazione TGM11

Si riporta l'output fornito da *Gretl*. Il modello di regressione stimato, in funzione dei coefficienti ottenuti, assume la forma:

$$\begin{aligned} \text{logit}[P(Y = 1)] \\ = -3.4006 - 1.2120x_1 - 0.0047x_2 + 0.0027x_3 + 0.0035x_4 + 0.0131x_5 \\ - 0.0001x_6 + 0.8123x_7 \end{aligned}$$

La probabilità di rottura stimata è calcolata come:

$$\hat{\pi}(x) = \frac{e^{-3.4006-1.2120x_1-0.0047x_2+0.0027x_3+0.0035x_4+0.0131x_5-0.0001x_6+0.8123x_7}}{1 + e^{-3.4006-1.2120x_1-0.0047x_2+0.0027x_3+0.0035x_4+0.0131x_5-0.0001x_6+0.8123x_7}}$$

```

Modello 1: Logit multinomiale, usando le osservazioni 1-420
Variabile dipendente: flag
Errori standard basati sull'Hessiana

      coefficiente      errore std.      z      p-value
-----
flag = 1
const      -3,40063      11,0588      -0,3075      0,7585
N_materiale -1,21204      0,242633      -4,995      5,87e-07 ***
diametro    -0,00473351    0,00249094      -1,900      0,0574 *
anno_posa    0,00277366      0,00589670      0,4704      0,6381
pMax         0,00355501    0,00772631      0,4601      0,6454
velocitaA    0,0131732      0,0339293      0,3883      0,6978
flusso       -0,000138098      0,000798257     -0,1730      0,8627
TramRoute    0,812319      0,390931      2,078      0,0377 **

Media var. dipendente  0,150000      SQM var. dipendente  0,357497
Log-verosimiglianza    -121,8626      Criterio di Akaike    259,7252
Criterio di Schwarz     292,0472      Hannan-Quinn          272,5003
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 372 (88,6%)
Test del rapporto di verosimiglianza: Chi-quadro(7) = 111,35 [0,0000]

```

Figura 6-11 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM11

Il segno algebrico dei coefficienti fornisce precisazioni, fondamentali e importanti per comprendere la funzionalità del modello:

- il segno negativo di β_1 indica che all'aumentare del numero assegnato a ciascun materiale, la probabilità di rottura diminuisce;
- il segno negativo di β_2 indica che all'aumentare del diametro della condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_3 indica che all'aumentare dell'anno di posa (quindi quanto più una condotta è recente), la probabilità di rottura aumenta. Condotte più giovani hanno maggiore probabilità di rottura;
- il segno positivo di β_4 indica che all'aumentare della pressione massima di esercizio della condotta, la probabilità di rottura aumenta;
- il segno positivo di β_5 indica che all'aumentare della velocità media dei carichi dinamici sovrastanti la condotta, la probabilità di rottura aumenta;
- il segno negativo di β_6 indica che all'aumentare del flusso veicolare, la probabilità di rottura diminuisce;
- il segno positivo di β_7 indica che la presenza del tram fa sì che la probabilità di rottura aumenti.

I valori del *Test di Z* relativi ad ogni variabile, fissato un livello di significatività α pari a 0.05, mostrano che le variabili: diametro, anno di posa, pressione massima, velocità media e flusso non risultano significative.

Inoltre il software mostra il valore della statistica *Chi-quadro* alla voce ‘*Test del rapporto verosimiglianze*’. Il valore restituito dal modello, con 7 gradi di libertà, riportati fra parentesi tonde, soddisfa la relazione:

$$G = 111.35 > \chi^2_{\frac{\alpha}{2}} = 12.38$$

Conseguentemente l’ipotesi nulla può essere rifiutata ed è possibile affermare che il modello completo aggiunge informazioni rispetto al modello con la sola intercetta e che la previsione viene implementata se viene presa in considerazione la variabile:

- materiale;
- tram route.

Estrazione TGM12

Si riporta l’output fornito da *Gretl*. Il modello di regressione stimato, in funzione dei coefficienti ottenuti, assume la forma:

$$\begin{aligned} \text{logit}[P(Y = 1)] \\ = -11.8319 - 1.1552x_1 - 0.0050x_2 + 0.0070x_3 + 0.0043x_4 \\ + 0.0013x_5 + 0.00004x_6 + 0.9584x_7 \end{aligned}$$

La probabilità di rottura stimata è calcolata come:

$$\hat{\pi}(x) = \frac{e^{-11.8319 - 1.1552x_1 - 0.0050x_2 + 0.0070x_3 + 0.0043x_4 + 0.0013x_5 + 0.00004x_6 + 0.9584x_7}}{1 + e^{-11.8319 - 1.1552x_1 - 0.0050x_2 + 0.0070x_3 + 0.0043x_4 + 0.0013x_5 + 0.00004x_6 + 0.9584x_7}}$$

Il segno algebrico dei coefficienti fornisce precisazioni, fondamentali e importanti per comprendere la funzionalità del modello:

- il segno negativo di β_1 indica che all’aumentare del numero assegnato a ciascun materiale, la probabilità di rottura diminuisce;
- il segno negativo di β_2 indica che all’aumentare del diametro della condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_3 indica che all’aumentare dell’anno di posa (quindi quanto più una condotta è recente), la probabilità di rottura aumenta. Condotte più giovani hanno maggiore probabilità di rottura;
- il segno positivo di β_4 indica che all’aumentare della pressione massima di esercizio della condotta, la probabilità di rottura aumenta;
- il segno positivo di β_5 indica che all’aumentare della velocità media dei carichi dinamici sovrastanti la condotta, la probabilità di rottura aumenta;

- il segno positivo di β_6 indica che all'aumentare del flusso veicolare, la probabilità di rottura aumenta;
- il segno positivo di β_7 indica che la presenza del tram fa sì che la probabilità di rottura aumenti.

```
Modello 1: Logit multinomiale, usando le osservazioni 1-420
Variabile dipendente: flag
Errori standard basati sull'Hessiana
```

	coefficiente	errore std.	z	p-value	

flag = 1					
const	-11,8319	10,8896	-1,087	0,2772	
N_materiale	-1,15520	0,245697	-4,702	2,58e-06	***
diametro	-0,00502855	0,00237370	-2,118	0,0341	**
anno_posa	0,00703036	0,00583272	1,205	0,2281	
pMax	0,00432040	0,00817607	0,5284	0,5972	
velocita	0,00133699	0,0382316	0,03497	0,9721	
flusso	4,06197e-05	0,000803817	0,05053	0,9597	
TramRoute	0,958439	0,356806	2,686	0,0072	***

Media var. dipendente	0,150000	SQM var. dipendente	0,357497		
Log-verosimiglianza	-125,8983	Criterio di Akaike	267,7967		
Criterio di Schwarz	300,1187	Hannan-Quinn	280,5718		
Note: SQM = scarto quadratico medio; E.S. = errore standard					
Numero dei casi 'previsti correttamente' = 375 (89,3%)					
Test del rapporto di verosimiglianza: Chi-quadro(7) = 103,279 [0,0000]					

Figura 6-12 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM12

I valori del *Test di Z* relativi ad ogni variabile, fissato un livello di significatività α pari a 0.05, mostrano che le variabili: anno di posa, pressione massima, velocità media e flusso non risultano significative.

Inoltre il software mostra il valore della statistica *Chi-quadro* alla voce '*Test del rapporto verosimiglianze*'. Il valore restituito dal modello, con 7 gradi di libertà, riportati fra parentesi tonde, soddisfa la relazione:

$$G = 103.279 > \chi^2_{\frac{\alpha}{2}} = 12.38$$

Conseguentemente l'ipotesi nulla può essere rifiutata ed è possibile affermare che il modello completo aggiunge informazioni rispetto al modello con la sola intercetta e che la previsione viene implementata se viene presa in considerazione la variabile:

- materiale;
- diametro;
- tram route.

Estrazione TGM13

L'output fornito da *Gretl* permette di definire il modello di regressione nella seguente forma:

$$\begin{aligned} \text{logit}[P(Y = 1)] \\ = -4.0269 - 1.2360x_1 - 0.0052x_2 + 0.0029x_3 - 0.0030x_4 + 0.0540x_5 \\ - 0.0005x_6 + 1.1713x_7 \end{aligned}$$

E la probabilità di rottura diviene:

$$\hat{\pi}(x) = \frac{e^{-4.0269-1.2360x_1-0.0052x_2+0.0029x_3-0.0030x_4+0.0540x_5-0.0005x_6+1.1713x_7}}{1 + e^{-4.0269-1.2360x_1-0.0052x_2+0.0029x_3-0.0030x_4+0.0540x_5-0.0005x_6+1.1713x_7}}$$

```
Modello 1: Logit multinomiale, usando le osservazioni 1-420
Variabile dipendente: flag
Errori standard basati sull'Hessiana
```

	coefficiente	errore std.	z	p-value
flag = 1				
const	-4,02695	10,8945	-0,3696	0,7117
N_materiale	-1,23606	0,245487	-5,035	4,77e-07 ***
diametro	-0,00529983	0,00235775	-2,248	0,0246 **
anno_posa	0,00297551	0,00581497	0,5117	0,6089
pMax	-0,00306504	0,00839043	-0,3653	0,7149
velocita	0,0540965	0,0412753	1,311	0,1900
flusso	-0,000592498	0,000828671	-0,7150	0,4746
TramRoute	1,17135	0,401635	2,916	0,0035 ***
Media var. dipendente	0,150000	SQM var. dipendente	0,357497	
Log-verosimiglianza	-117,4295	Criterio di Akaike	250,8589	
Criterio di Schwarz	283,1810	Hannan-Quinn	263,6341	

Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 380 (90,5%)
 Test del rapporto di verosimiglianza: Chi-quadro(7) = 120,217 [0,0000]

Figura 6-13 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM13

Si traggono le seguenti considerazioni:

- il segno negativo di β_1 indica che all'aumentare del numero assegnato a ciascun materiale, la probabilità di rottura diminuisce;
- il segno negativo di β_2 indica che all'aumentare del diametro della condotta, la probabilità di rottura diminuisce;

- il segno positivo di β_3 indica che all'aumentare dell'anno di posa (quindi quanto più una condotta è recente), la probabilità di rottura aumenta. Condotte più giovani hanno maggiore probabilità di rottura;
- il segno negativo di β_4 indica che all'aumentare della pressione massima di esercizio della condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_5 indica che all'aumentare della velocità media dei carichi dinamici sovrastanti la condotta, la probabilità di rottura aumenta;
- il segno negativo di β_6 indica che all'aumentare del flusso veicolare, la probabilità di rottura diminuisce;
- il segno positivo di β_7 indica che la presenza del tram fa sì che la probabilità di rottura aumenti.

I valori del *Test di Z* relativi ad ogni variabile, fissato un livello di significatività α pari a 0.05, mostrano che variabili quali: anno di posa, pressione massima, velocità media e flusso non risultano significative.

Inoltre il software mostra il valore della statistica *Chi-quadro* alla voce '*Test del rapporto verosimiglianze*'. Il valore restituito dal modello, con 7 gradi di libertà, riportati fra parentesi tonde, soddisfa la relazione:

$$G = 120.217 > \chi^2_{\frac{\alpha}{2}} = 12.38$$

Conseguentemente l'ipotesi nulla può essere rifiutata ed è possibile affermare che il modello completo aggiunge informazioni rispetto al modello con la sola intercetta e che la previsione viene implementata se vengono prese in considerazione le variabili:

- materiale;
- diametro;
- tram route.

Estrazione TGM14

Si riporta l'output fornito da *Gretl*. Il modello di regressione stimato, in funzione dei coefficienti ottenuti, assume la forma:

$$\begin{aligned} \text{logit}[P(Y = 1)] \\ = -6.8429 - 1.1656x_1 - 0.0053x_2 + 0.0037x_3 + 0.0341x_4 + 0.0065x_5 \\ + 0.00004x_6 + 0.9238x_7 \end{aligned}$$

La probabilità di rottura stimata è calcolata come:

$$\hat{\pi}(x) = \frac{e^{-6.8429-1.1656x_1-0.0053x_2+0.0037x_3+0.0341x_4+0.0065x_5+0.00004x_6+0.9238x_7}}{1 + e^{-6.8429-1.1656x_1-0.0053x_2+0.0037x_3+0.0341x_4+0.0065x_5+0.00004x_6+0.9238x_7}}$$

Modello 1: Logit multinomiale, usando le osservazioni 1-420

Variabile dipendente: flag

Errori standard basati sull'Hessiana

	coefficiente	errore std.	z	p-value	

flag = 1					
const	-6,84298	11,2006	-0,6109	0,5412	
N_materiale	-1,16562	0,238700	-4,883	1,04e-06	***
diametro	-0,00534191	0,00231334	-2,309	0,0209	**
anno_posa	0,00374802	0,00596216	0,6286	0,5296	
pMax	0,0341839	0,0190539	1,794	0,0728	*
velocitA	0,00659622	0,0368006	0,1792	0,8577	
flusso	4,31756e-05	0,000854290	0,05054	0,9597	
TramRoute	0,923855	0,391273	2,361	0,0182	**
Media var. dipendente	0,150000	SQM var. dipendente	0,357497		
Log-verosimiglianza	-119,7634	Criterio di Akaike	255,5269		
Criterio di Schwarz	287,8489	Hannan-Quinn	268,3020		
Note: SQM = scarto quadratico medio; E.S. = errore standard					
Numero dei casi 'previsti correttamente' = 375 (89,3%)					
Test del rapporto di verosimiglianza: Chi-quadro(7) = 115,549 [0,0000]					

Figura 6-14 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM14

Il segno algebrico dei coefficienti fornisce precisazioni, fondamentali e importanti per comprendere la funzionalità del modello:

- il segno negativo di β_1 indica che all'aumentare del numero assegnato a ciascun materiale, la probabilità di rottura diminuisce;
- il segno negativo di β_2 indica che all'aumentare del diametro della condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_3 indica che all'aumentare dell'anno di posa (quindi quanto più una condotta è recente), la probabilità di rottura aumenta. Condotte più giovani hanno maggiore probabilità di rottura;
- il segno positivo di β_4 indica che all'aumentare della pressione massima di esercizio della condotta, la probabilità di rottura aumenta;
- il segno positivo di β_5 indica che all'aumentare della velocità media dei carichi dinamici sovrastanti la condotta, la probabilità di rottura aumenta;
- il segno positiva di β_6 indica che all'aumentare del flusso veicolare, la probabilità di rottura aumenta;
- il segno positivo di β_7 indica che la presenza del tram fa sì che la probabilità di rottura aumenti.

I valori del *Test di Z* relativi ad ogni variabile, fissato un livello di significatività α pari a 0.05, mostrano che le variabili: anno di posa, pressione massima, velocità media e flusso non risultano significative.

Inoltre il software mostra il valore della statistica *Chi-quadro* alla voce '*Test del rapporto verosimiglianze*'. Il valore restituito dal modello, con 7 gradi di libertà, riportati fra parentesi tonde, soddisfa la relazione:

$$G = 115.549 > \chi^2_{\frac{\alpha}{2}} = 12.38$$

Conseguentemente l'ipotesi nulla può essere rifiutata ed è possibile affermare che il modello completo aggiunge informazioni rispetto al modello con la sola intercetta e che la previsione viene implementata se vengono prese in considerazione le variabili:

- materiale;
- diametro;
- tram route.

Estrazione TGM15

L'output fornito da *Gretl* permette di definire il modello di regressione nella seguente forma:

$$\begin{aligned} \text{logit}[P(Y = 1)] \\ = -17.8783 - 1.3654x_1 - 0.0058x_2 + 0.0102x_3 + 0.0102x_4 \\ + 0.0187x_5 + 0.00007x_6 + 1.4192x_7 \end{aligned}$$

E la probabilità di rottura diviene:

$$\hat{\pi}(x) = \frac{e^{-17.8783-1.3654x_1-0.0058x_2+0.0102x_3+0.0102x_4+0.0187x_5+0.00007x_6+1.4192x_7}}{1 + e^{-17.8783-1.3654x_1-0.0058x_2+0.0102x_3+0.0102x_4+0.0187x_5+0.00007x_6+1.4192x_7}}$$

Si traggono le seguenti considerazioni:

- il segno negativo di β_1 indica che all'aumentare del numero assegnato a ciascun materiale, la probabilità di rottura diminuisce;
- il segno negativo di β_2 indica che all'aumentare del diametro della condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_3 indica che all'aumentare dell'anno di posa (quindi quanto più una condotta è recente), la probabilità di rottura aumenta. Condotte più giovani hanno maggiore probabilità di rottura;

- il segno positivo di β_4 indica che all'aumentare della pressione massima di esercizio della condotta, la probabilità di rottura aumenta;
- il segno positivo di β_5 indica che all'aumentare della velocità media dei carichi dinamici sovrastanti la condotta, la probabilità di rottura aumenta;
- il segno positivo di β_6 indica che all'aumentare del flusso veicolare, la probabilità di rottura aumenta;
- il segno positivo di β_7 indica che la presenza del tram fa sì che la probabilità di rottura aumenti.

```
Modello 1: Logit multinomiale, usando le osservazioni 1-420
Variabile dipendente: flag
Errori standard basati sull'Hessiana
```

	coefficiente	errore std.	z	p-value	

flag = 1					
const	-17,8783	11,6479	-1,535	0,1248	
N_materiale	-1,36540	0,254532	-5,364	8,12e-08	***
diametro	-0,00587326	0,00231494	-2,537	0,0112	**
anno_posa	0,0101209	0,00622603	1,626	0,1040	
pMax	0,0102350	0,0109579	0,9340	0,3503	
velocità	0,0187330	0,0380629	0,4922	0,6226	
flusso	7,15995e-05	0,000832271	0,08603	0,9314	
TramRoute	1,41925	0,399865	3,549	0,0004	***

Media var. dipendente	0,150000	SQM var. dipendente	0,357497		
Log-verosimiglianza	-115,8373	Criterio di Akaike	247,6746		
Criterio di Schwarz	279,9966	Hannan-Quinn	260,4498		
Note: SQM = scarto quadratico medio; E.S. = errore standard					
Numero dei casi 'previsti correttamente' = 383 (91,2%)					
Test del rapporto di verosimiglianza: Chi-quadro(7) = 123,401 [0,0000]					

Figura 6-15 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM15

I valori del *Test di Z* relativi ad ogni variabile, fissato un livello di significatività α pari a 0.05, mostrano che variabili quali: anno di posa, pressione massima, velocità media e flusso non risultano significative.

Inoltre il software mostra il valore della statistica *Chi-quadro* alla voce '*Test del rapporto verosimiglianze*'. Il valore restituito dal modello, con 7 gradi di libertà, riportati fra parentesi tonde, soddisfa la relazione:

$$G = 123.401 > \chi_{\frac{\alpha}{2}}^2 = 12.38$$

Conseguentemente l'ipotesi nulla può essere rifiutata ed è possibile affermare che il modello completo aggiunge informazioni rispetto al modello con la sola intercetta e che la previsione viene implementata se vengono prese in considerazione le variabili:

- materiale;
- diametro;
- tram route.

Estrazione TGM16

Si riporta l'output fornito da *Gretl*. Il modello di regressione stimato, in funzione dei coefficienti ottenuti, assume la forma:

$$\begin{aligned} \text{logit}[P(Y = 1)] \\ = -4.3757 - 1.2218x_1 - 0.0053x_2 + 0.0030x_3 + 0.0032x_4 + 0.0315x_5 \\ + 0.0003x_6 + 0.7595x_7 \end{aligned}$$

La probabilità di rottura stimata è calcolata come:

$$\hat{\pi}(x) = \frac{e^{-4.3757-1.2218x_1-0.0053x_2+0.0030x_3+0.0032x_4+0.0315x_5+0.0003x_6+0.7595x_7}}{1 + e^{-4.3757-1.2218x_1-0.0053x_2+0.0030x_3+0.0032x_4+0.0315x_5+0.0003x_6+0.7595x_7}}$$

Modello 1: Logit multinomiale, usando le osservazioni 1-420				
Variabile dipendente: flag				
Errori standard basati sull'Hessiana				
	coefficiente	errore std.	z	p-value

flag = 1				
const	-4,37573	11,4366	-0,3826	0,7020
N_materiale	-1,22184	0,251146	-4,865	1,14e-06 ***
diámetro	-0,00534979	0,00241746	-2,213	0,0269 **
anno_posa	0,00308433	0,00607719	0,5075	0,6118
pMax	0,00321224	0,00769659	0,4174	0,6764
velocità	0,0315271	0,0363339	0,8677	0,3856
flusso	0,000306852	0,000750346	0,4089	0,6826
TramRoute	0,759520	0,392425	1,935	0,0529 *
Media var. dipendente	0,150000	SQM var. dipendente	0,357497	
Log-verosimiglianza	-119,6323	Criterio di Akaike	255,2646	
Criterio di Schwarz	287,5866	Hannan-Quinn	268,0397	
Note: SQM = scarto quadratico medio; E.S. = errore standard				
Numero dei casi 'previsti correttamente' = 370 (88,1%)				
Test del rapporto di verosimiglianza: Chi-quadro(7) = 115,811 [0,0000]				

Figura 6-16 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM16

Il segno algebrico dei coefficienti fornisce precisazioni, fondamentali e importanti per comprendere la funzionalità del modello:

- il segno negativo di β_1 indica che all'aumentare del numero assegnato a ciascun materiale, la probabilità di rottura diminuisce;
- il segno negativo di β_2 indica che all'aumentare del diametro della condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_3 indica che all'aumentare dell'anno di posa (quindi quanto più una condotta è recente), la probabilità di rottura aumenta. Condotte più giovani hanno maggiore probabilità di rottura;
- il segno positivo di β_4 indica che all'aumentare della pressione massima di esercizio della condotta, la probabilità di rottura aumenta;
- il segno positivo di β_5 indica che all'aumentare della velocità media dei carichi dinamici sovrastanti la condotta, la probabilità di rottura aumenta;
- il segno positivo di β_6 indica che all'aumentare del flusso veicolare, la probabilità di rottura aumenta;
- il segno positivo di β_7 indica che la presenza del tram fa sì che la probabilità di rottura aumenti.

I valori del *Test di Z* relativi ad ogni variabile, fissato un livello di significatività α pari a 0.05, mostrano che le variabili: anno di posa, pressione massima, velocità media, flusso e tram non risultano significative.

Inoltre il software mostra il valore della statistica *Chi-quadro* alla voce '*Test del rapporto verosimiglianze*'. Il valore restituito dal modello, con 7 gradi di libertà, riportati fra parentesi tonde, soddisfa la relazione:

$$G = 115.811 > \chi^2_{\frac{\alpha}{2}} = 12.38$$

Conseguentemente l'ipotesi nulla può essere rifiutata ed è possibile affermare che il modello completo aggiunge informazioni rispetto al modello con la sola intercetta e che la previsione viene implementata se viene presa in considerazione la variabile:

- materiale;
- diametro.

Estrazione TGM17

Si riporta l'output fornito da *Gretl*. Il modello di regressione stimato, in funzione dei coefficienti ottenuti, assume la forma:

$$\begin{aligned} \text{logit}[P(Y = 1)] \\ = -11.2381 - 1.1390x_1 - 0.0055x_2 + 0.0062x_3 + 0.0254x_4 \\ + 0.0009x_5 - 0.0001x_6 + 0.8404x_7 \end{aligned}$$

La probabilità di rottura stimata è calcolata come:

$$\hat{\pi}(x) = \frac{e^{-11.2381-1.1390x_1-0.0055x_2+0.0062x_3+0.0254x_4+0.0009x_5-0.0001x_6+0.8404x_7}}{1 + e^{-11.2381-1.1390x_1-0.0055x_2+0.0062x_3+0.0254x_4+0.0009x_5-0.0001x_6+0.8404x_7}}$$

Modello 1: Logit multinomiale, usando le osservazioni 1-420
 Variabile dipendente: flag
 Errori standard basati sull'Hessiana

	coefficiente	errore std.	z	p-value	
flag = 1					
const	-11,2381	10,8394	-1,037	0,2998	
N_materiale	-1,13903	0,237032	-4,805	1,54e-06	***
diametro	-0,00552133	0,00222463	-2,482	0,0131	**
anno_posa	0,00624246	0,00580538	1,075	0,2822	
pMax	0,0254979	0,0186786	1,365	0,1722	
velocita	0,000962924	0,0351797	0,02737	0,9782	
flusso	-0,000144903	0,000779766	-0,1858	0,8526	
TramRoute	0,840422	0,370818	2,266	0,0234	**
Media var. dipendente	0,150000	SQM var. dipendente	0,357497		
Log-verosimiglianza	-126,9503	Criterio di Akaike	269,9006		
Criterio di Schwarz	302,2226	Hannan-Quinn	282,6757		

Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 368 (87,6%)
 Test del rapporto di verosimiglianza: Chi-quadro(7) = 101,175 [0,0000]

Figura 6-17 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM17

Il segno algebrico dei coefficienti fornisce precisazioni, fondamentali e importanti per comprendere la funzionalità del modello:

- il segno negativo di β_1 indica che all'aumentare del numero assegnato a ciascun materiale, la probabilità di rottura diminuisce;
- il segno negativo di β_2 indica che all'aumentare del diametro della condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_3 indica che all'aumentare dell'anno di posa (quindi quanto più una condotta è recente), la probabilità di rottura aumenta. Condotte più giovani hanno maggiore probabilità di rottura;
- il segno positivo di β_4 indica che all'aumentare della pressione massima di esercizio della condotta, la probabilità di rottura aumenta;
- il segno positivo di β_5 indica che all'aumentare della velocità media dei carichi dinamici sovrastanti la condotta, la probabilità di rottura aumenta;
- il segno negativo di β_6 indica che all'aumentare del flusso veicolare, la probabilità di rottura diminuisce;

- il segno positivo di β_7 indica che la presenza del tram fa sì che la probabilità di rottura aumenti.

I valori del *Test di Z* relativi ad ogni variabile, fissato un livello di significatività α pari a 0.05, mostrano che le variabili: anno di posa, pressione massima, velocità media e flusso non risultano significative.

Inoltre il software mostra il valore della statistica *Chi-quadro* alla voce '*Test del rapporto verosimiglianze*'. Il valore restituito dal modello, con 7 gradi di libertà, riportati fra parentesi tonde, soddisfa la relazione:

$$G = 101.175 > \chi^2_{\frac{\alpha}{2}} = 12.38$$

Conseguentemente l'ipotesi nulla può essere rifiutata ed è possibile affermare che il modello completo aggiunge informazioni rispetto al modello con la sola intercetta e che la previsione viene implementata se viene presa in considerazione la variabile:

- materiale;
- diametro;
- tram route.

Estrazione TGM18

Si riporta l'output fornito da *Gretl*. Il modello di regressione stimato, in funzione dei coefficienti ottenuti, assume la forma:

$$\begin{aligned} \text{logit}[P(Y = 1)] \\ = -9.3148 - 1.3325x_1 - 0.0054x_2 + 0.0060x_3 + 0.0013x_4 + 0.0035x_5 \\ + 0.0003x_6 + 0.8145x_7 \end{aligned}$$

La probabilità di rottura stimata è calcolata come:

$$\hat{\pi}(x) = \frac{e^{-9.3148 - 1.3325x_1 - 0.0054x_2 + 0.0060x_3 + 0.0013x_4 + 0.0035x_5 + 0.0003x_6 + 0.8145x_7}}{1 + e^{-9.3148 - 1.3325x_1 - 0.0054x_2 + 0.0060x_3 + 0.0013x_4 + 0.0035x_5 + 0.0003x_6 + 0.8145x_7}}$$

Il segno algebrico dei coefficienti fornisce precisazioni, fondamentali e importanti per comprendere la funzionalità del modello:

- il segno negativo di β_1 indica che all'aumentare del numero assegnato a ciascun materiale, la probabilità di rottura diminuisce;
- il segno negativo di β_2 indica che all'aumentare del diametro della condotta, la probabilità di rottura diminuisce;

- il segno positivo di β_3 indica che all'aumentare dell'anno di posa (quindi quanto più una condotta è recente), la probabilità di rottura aumenta. Condotte più giovani hanno maggiore probabilità di rottura;
- il segno positivo di β_4 indica che all'aumentare della pressione massima di esercizio della condotta, la probabilità di rottura aumenta;
- il segno positivo di β_5 indica che all'aumentare della velocità media dei carichi dinamici sovrastanti la condotta, la probabilità di rottura aumenta;
- il segno positivo di β_6 indica che all'aumentare del flusso veicolare, la probabilità di rottura aumenta;
- il segno positivo di β_7 indica che la presenza del tram fa sì che la probabilità di rottura aumenti.

Modello 1: Logit multinomiale, usando le osservazioni 1-420

Variabile dipendente: flag

Errori standard basati sull'Hessiana

	coefficiente	errore std.	z	p-value	

flag = 1					
const	-9,31483	11,9670	-0,7784	0,4363	
N_materiale	-1,33259	0,262112	-5,084	3,69e-07	***
diametro	-0,00547324	0,00246189	-2,223	0,0262	**
anno_posa	0,00609115	0,00639687	0,9522	0,3410	
pMax	0,00137343	0,00858113	0,1601	0,8728	
velocità	0,00351950	0,0367318	0,09582	0,9237	
flusso	0,000351423	0,000821914	0,4276	0,6690	
TramRoute	0,814512	0,384027	2,121	0,0339	**
Media var. dipendente	0,150000	SQM var. dipendente		0,357497	
Log-verosimiglianza	-118,9091	Criterio di Akaike		253,8182	
Criterio di Schwarz	286,1403	Hannan-Quinn		266,5934	
Note: SQM = scarto quadratico medio; E.S. = errore standard					
Numero dei casi 'previsti correttamente' = 376 (89,5%)					
Test del rapporto di verosimiglianza: Chi-quadro(7) = 117,257 [0,0000]					

Figura 6-18 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM18

I valori del *Test di Z* relativi ad ogni variabile, fissato un livello di significatività α pari a 0.05, mostrano che le variabili: anno di posa, pressione massima, velocità media e flusso non risultano significative.

Inoltre il software mostra il valore della statistica *Chi-quadro* alla voce '*Test del rapporto verosimiglianze*'. Il valore restituito dal modello, con 7 gradi di libertà, riportati fra parentesi tonde, soddisfa la relazione:

$$G = 117.257 > \chi_{\frac{\alpha}{2}}^2 = 12.38$$

Conseguentemente l'ipotesi nulla può essere rifiutata ed è possibile affermare che il modello completo aggiunge informazioni rispetto al modello con la sola intercetta e che la previsione viene implementata se viene presa in considerazione la variabile:

- materiale;
- diametro;
- tram route.

Estrazione TGM19

Si riporta l'output fornito da *Gretl*. Il modello di regressione stimato, in funzione dei coefficienti ottenuti, assume la forma:

$$\begin{aligned} \text{logit}[P(Y = 1)] \\ = 0.6505 - 1.0814x_1 - 0.0054x_2 + 0.0005x_3 + 0.0011x_4 + 0.0181x_5 \\ - 0.00007x_6 + 0.8941x_7 \end{aligned}$$

La probabilità di rottura stimata è calcolata come:

$$\hat{\pi}(x) = \frac{e^{0.6505 - 1.0814x_1 - 0.0054x_2 + 0.0005x_3 + 0.0011x_4 + 0.0181x_5 - 0.00007x_6 + 0.8941x_7}}{1 + e^{0.6505 - 1.0814x_1 - 0.0054x_2 + 0.0005x_3 + 0.0011x_4 + 0.0181x_5 - 0.00007x_6 + 0.8941x_7}}$$

Modello 1: Logit multinomiale, usando le osservazioni 1-420

Variabile dipendente: flag

Errori standard basati sull'Hessiana

	coefficiente	errore std.	z	p-value	

flag = 1					
const	0,650544	11,0156	0,05906	0,9529	
N_materiale	-1,08149	0,238012	-4,544	5,52e-06	***
diámetro	-0,00545566	0,00243084	-2,244	0,0248	**
anno_posa	0,000571812	0,00587057	0,09740	0,9224	
pMax	0,00112529	0,00922213	0,1220	0,9029	
velocitA	0,0181410	0,0356428	0,5090	0,6108	
flusso	-7,90681e-05	0,000791527	-0,09989	0,9204	
TramRoute	0,894135	0,376573	2,374	0,0176	**
Media var. dipendente	0,150000	SQM var. dipendente	0,357497		
Log-verosimiglianza	-120,3421	Criterio di Akaike	256,6843		
Criterio di Schwarz	289,0063	Hannan-Quinn	269,4594		
Note: SQM = scarto quadratico medio; E.S. = errore standard					
Numero dei casi 'previsti correttamente' = 372 (88,6%)					
Test del rapporto di verosimiglianza: Chi-quadro(7) = 114,391 [0,0000]					

Figura 6-19 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM19

Il segno algebrico dei coefficienti fornisce precisazioni, fondamentali e importanti per comprendere la funzionalità del modello:

- il segno negativo di β_1 indica che all'aumentare del numero assegnato a ciascun materiale, la probabilità di rottura diminuisce;
- il segno negativo di β_2 indica che all'aumentare del diametro della condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_3 indica che all'aumentare dell'anno di posa (quindi quanto più una condotta è recente), la probabilità di rottura aumenta. Condotte più giovani hanno maggiore probabilità di rottura;
- il segno positivo di β_4 indica che all'aumentare della pressione massima di esercizio della condotta, la probabilità di rottura aumenta;
- il segno positivo di β_5 indica che all'aumentare della velocità media dei carichi dinamici sovrastanti la condotta, la probabilità di rottura aumenta;
- il segno negativo di β_6 indica che all'aumentare del flusso veicolare, la probabilità di rottura diminuisce;
- il segno positivo di β_7 indica che la presenza del tram fa sì che la probabilità di rottura aumenti.

I valori del *Test di Z* relativi ad ogni variabile, fissato un livello di significatività α pari a 0.05, mostrano che le variabili: anno di posa, pressione massima, velocità media e flusso non risultano significative.

Inoltre il software mostra il valore della statistica *Chi-quadro* alla voce '*Test del rapporto verosimiglianze*'. Il valore restituito dal modello, con 7 gradi di libertà, riportati fra parentesi tonde, soddisfa la relazione:

$$G = 114.391 > \chi^2_{\alpha/2} = 12.38$$

Conseguentemente l'ipotesi nulla può essere rifiutata ed è possibile affermare che il modello completo aggiunge informazioni rispetto al modello con la sola intercetta e che la previsione viene implementata se viene presa in considerazione la variabile:

- materiale;
- diametro;
- tram route.

Estrazione TGM20

Si riporta l'output fornito da *Gretl*. Il modello di regressione stimato, in funzione dei coefficienti ottenuti, assume la forma:

$$\begin{aligned} \text{logit}[P(Y = 1)] \\ = -6.0725 - 1.1307x_1 - 0.0047x_2 + 0.0036x_3 + 0.0267x_4 - 0.0210x_5 \\ + 0.0005x_6 + 0.7360x_7 \end{aligned}$$

La probabilità di rottura stimata è calcolata come:

$$\hat{\pi}(x) = \frac{e^{-6.0725-1.1307x_1-0.0047x_2+0.0036x_3+0.0267x_4-0.0210x_5+0.0005x_6+0.7360x_7}}{1 + e^{-6.0725-1.1307x_1-0.0047x_2+0.0036x_3+0.0267x_4-0.0210x_5+0.0005x_6+0.7360x_7}}$$

Modello 1: Logit multinomiale, usando le osservazioni 1-420				
Variabile dipendente: flag				
Errori standard basati sull'Hessiana				
	coefficiente	errore std.	z	p-value

flag = 1				
const	-6,07256	10,8367	-0,5604	0,5752
N_materiale	-1,13075	0,234816	-4,815	1,47e-06 ***
diametro	-0,00477831	0,00219610	-2,176	0,0296 **
anno_posa	0,00361991	0,00580323	0,6238	0,5328
pMax	0,0267867	0,0140043	1,913	0,0558 *
velocita	-0,0210808	0,0356838	-0,5908	0,5547
flusso	0,000523796	0,000818965	0,6396	0,5224
TramRoute	0,736021	0,381148	1,931	0,0535 *
Media var. dipendente	0,150000	SQM var. dipendente	0,357497	
Log-verosimiglianza	-125,8925	Criterio di Akaike	267,7850	
Criterio di Schwarz	300,1071	Hannan-Quinn	280,5602	
Note: SQM = scarto quadratico medio; E.S. = errore standard				
Numero dei casi 'previsti correttamente' = 369 (87,9%)				
Test del rapporto di verosimiglianza: Chi-quadro(7) = 103,291 [0,0000]				

Figura 6-20 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM20

Il segno algebrico dei coefficienti fornisce precisazioni, fondamentali e importanti per comprendere la funzionalità del modello:

- il segno negativo di β_1 indica che all'aumentare del numero assegnato a ciascun materiale, la probabilità di rottura diminuisce;
- il segno negativo di β_2 indica che all'aumentare del diametro della condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_3 indica che all'aumentare dell'anno di posa (quindi quanto più una condotta è recente), la probabilità di rottura aumenta. Condotte più giovani hanno maggiore probabilità di rottura;
- il segno positivo di β_4 indica che all'aumentare della pressione massima di esercizio della condotta, la probabilità di rottura aumenta;

- il segno negativo di β_5 indica che all'aumentare della velocità media dei carichi dinamici sovrastanti la condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_6 indica che all'aumentare del flusso veicolare, la probabilità di rottura aumenta;
- il segno positivo di β_7 indica che la presenza del tram fa sì che la probabilità di rottura aumenti.

I valori del *Test di Z* relativi ad ogni variabile, fissato un livello di significatività α pari a 0.05, mostrano che le variabili: anno di posa, pressione massima, velocità media, flusso e tram non risultano significative.

Inoltre il software mostra il valore della statistica *Chi-quadro* alla voce '*Test del rapporto verosimiglianze*'. Il valore restituito dal modello, con 7 gradi di libertà, riportati fra parentesi tonde, soddisfa la relazione:

$$G = 103.291 > \chi^2_{\frac{\alpha}{2}} = 12.38$$

Conseguentemente l'ipotesi nulla può essere rifiutata ed è possibile affermare che il modello completo aggiunge informazioni rispetto al modello con la sola intercetta e che la previsione viene implementata se viene presa in considerazione la variabile:

- materiale;
- diametro.

Osservazioni

Dall'analisi dei 20 set di dati, estratti dalla tabella *RegressioneTGM* in modo da avere una percentuale di rottura pari alla tabella *ReteTorino*, si evince che:

- le variabili *pressione massima*, *velocità media* e *flusso veicolare* risultano sempre poco significative, come dimostrano i valori del *p-value* ad esse associati;
- in tutte le estrazioni la variabile *materiale* risulta rilevante per il modello;
- la variabile *diametro* non risulta significativa in 4 estrazioni, mentre *Tram Route* non è una variabile rilevante per il modello in 5 estrazioni su 20;
- in una sola estrazione la variabile *anno di posa* risulta significativa;
- in tutte le estrazioni il coefficiente β della variabile *tram route* ha segno positivo: la probabilità di rottura aumenta se la variabile ha valore pari ad uno, indice della presenza del tram;
- in tutte le estrazioni i coefficienti β delle variabili *materiale* e *diametro* hanno segno negativo: la probabilità di rottura diminuisce all'aumentare dei valori che le variabili assumono.

6.2 VERIFICA DEI MODELLI APPLICANDO IL METODO BACKWARD

Per comprendere le prestazioni del modello, ognuno dei 20 set di dati è diviso in un set di allenamento, sul quale viene tarato il modello, e un set di test, sul quale avviene la previsione del modello. Un buon modello, per potersi ritenere tale, dovrebbe prevedere una probabilità di rottura maggiore per le condotte, che hanno già presentato rottura in passato, caratterizzate da flag pari a 1.

In questo caso, ogni set di dati è stato suddiviso in due parti in un rapporto 80:20, ciò significa che l'80% dei dati, ossia 336 condotte selezionate in maniera random, verrà utilizzato per l'allenamento del modello e il 20%, ossia le restanti 84 condotte di ogni *Estrazione_TGMn*, per il test del modello. Ogni nuovo campione ha una percentuale di rottura diversa, ottenuta dalla selezione casuale delle condotte.

La creazione del modello avviene applicando il metodo *backward*, che inizialmente prende in considerazione tutte e sette le variabili e successivamente le elimina una alla volta partendo da quella con l'associazione con la variabile dipendente meno significativa, valutata attraverso il *test di Z* e il valore del *p-value*.

Si descrive il procedimento, riportando nel dettaglio i passi eseguiti esclusivamente su uno dei set di fitting:

Estrazione TGM1-80%

Il nuovo set di dati è costituito dall'80% delle condotte presenti in '*Estrazione_TGM1*', su questo vengono stimati i parametri di regressione logistica, inserendo, dapprima, nel modello tutte e sette le variabili indipendenti. Si riporta l'output fornito dal software *Gretl*:

```
Modello 1: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana
```

	coefficiente	errore std.	z	p-value
flag = 1				
const	-21,3857	13,1684	-1,624	0,1044
N_materiale	-1,41372	0,281472	-5,023	5,10e-07 ***
diametro	-0,00641674	0,00280635	-2,287	0,0222 **
anno_posa	0,0118641	0,00709513	1,672	0,0945 *
pMax	0,0314651	0,0204336	1,540	0,1236
velocita	0,00781811	0,0393817	0,1985	0,8426
flusso	-0,00121828	0,000931930	-1,307	0,1911
TramRoute	1,30390	0,444829	2,931	0,0034 ***
Media var. dipendente	0,172619	SQM var. dipendente	0,378481	
Log-verosimiglianza	-101,4061	Criterio di Akaike	218,8121	
Criterio di Schwarz	249,3490	Hannan-Quinn	230,9850	

Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 298 (88,7%)
 Test del rapporto di verosimiglianza: Chi-quadro(7) = 106,318 [0,0000]

Figura 6-21 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM1-80%

L'indicatore visivo, così come i valori del *p-value*, segnalano la presenza di variabili poco significative. L'applicazione del metodo *backward* fa sì che i coefficienti di regressione vengano stimati con le sole variabili: *materiale*, *diametro* e *tram*, ritenute significative.

```

Eliminazione sequenziale usando alfa (a due code) = 0,05

Scartata velocitA          (p-value 0,843)
Scartata flusso           (p-value 0,164)
Scartata anno_posa        (p-value 0,147)
Scartata pMax             (p-value 0,082)

Test sul modello 1

Ipotesi nulla: i parametri della regressione valgono zero per le variabili
anno_posa, pMax, velocitA, flusso
Statistica test: F(4, 328) = 1,61387, p-value 0,170407
L'omissione delle variabili ha migliorato 3 dei 3 criteri di informazione.

Modello 2: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

-----
              coefficiente      errore std.      z      p-value
-----
flag = 1
const          2,14566          0,569826          3,765    0,0002 ***
N_materiale    -1,02739          0,175798          -5,844    5,09e-09 ***
diametro       -0,00620779       0,00301104        -2,062    0,0392 **
TramRoute       0,853462          0,360544          2,367    0,0179 **

Media var. dipendente  0,172619      SQM var. dipendente  0,378481
Log-verosimiglianza   -104,9345      Criterio di Akaike    217,8690
Criterio di Schwarz    233,1375      Hannan-Quinn          223,9555
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 293 (87,2%)
Test del rapporto di verosimiglianza: Chi-quadro(3) = 99,2609 [0,0000]

```

Figura 6-22 Stime dei parametri del modello applicato al set di dati Estrazione_TGMI-80% con variabili significative

Il modello di regressione logistica multivariata, in relazione ai risultati ottenuti, assume la forma:

$$\text{logit}[P(Y = 1)] = 2.1456 - 1.0273x_1 - 0.0062x_2 + 0.8534x_7$$

Ottenuti i coefficienti di regressione, la fase di test del modello viene eseguita sul restante 20% dei tratti.

La probabilità di rottura per la generica condotta presente nel set di dati 'Estrazione_TGMI-20%' viene calcolata come:

$$\hat{\pi}(x) = \frac{e^{2.1456 - 1.0273x_1 - 0.0062x_2 + 0.8534x_7}}{1 + e^{2.1456 - 1.0273x_1 - 0.0062x_2 + 0.8534x_7}}$$

Per verificare la capacità del modello di predire una rottura, è stato creato un grafico a bolle, avente sull'asse delle ascisse i valori del flag, 0 ed 1, e sull'asse delle ordinate i valori delle probabilità stimate.

Le dimensioni di ogni bolla sono proporzionali all'occorrenza, cioè il numero di volte che una determinata percentuale di probabilità si ripete nella classe di condotte con flag pari a 0, in questo caso 79, e con flag pari a 1, in questo caso 5 condotte.

Inoltre per ogni classe è stato calcolato il valore di probabilità medio di rottura, riportato sul grafico con le bolle di colore arancione.

I valori medi confermano che la percentuale di rottura più alta è stimata sui tratti che hanno realmente presentato rottura nel decennio in analisi.

È quindi possibile affermare che il modello ottenuto è in grado di predire una rottura futura.

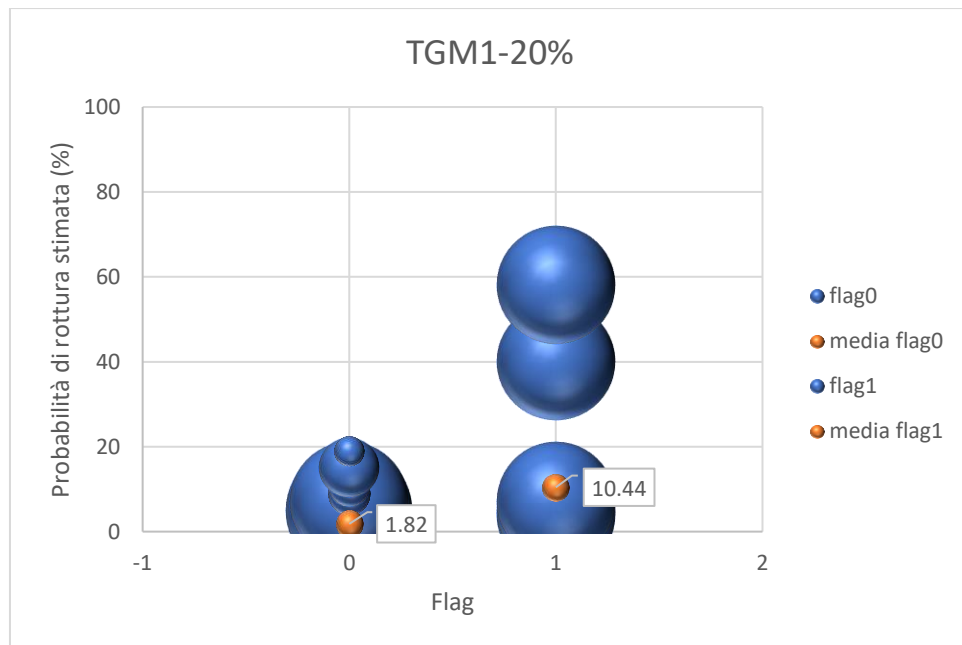


Figura 6-23 Confronto tra la probabilità di rottura stimata e il dato reale delle condotte dell'Estrazione_TGM1-20%

Tale tipo di procedura è stata eseguita per le restanti 19 estrazioni. In Tabella 6-2 si riportano i valori stimati da *Gretl* dei coefficienti di regressione, applicando il metodo *backward*, all'80% delle estrazioni. Le variabili non significative vengono indicate in tabella con 'n.s.' e vengono scartate dall'analisi.

Tabella 6-2 Coefficienti di regressione stimati sull'80% delle 20 estrazioni

Variabile	Materiale	Diametro	Anno di posa	p_Max	Velocità	Flusso veic.	TramRoute
Estrazione \ Coeff. Regressione	β_1	β_2	β_3	β_4	β_5	β_6	β_7
1	-1.02739	-0.0062	n.s.	n.s.	n.s.	n.s.	0.85346
2	-1.45405	-0.00534	0.01646	n.s.	n.s.	n.s.	1.01474
3	-1.0336	-0.00555	n.s.	n.s.	n.s.	n.s.	1.19311
4	-1.2281	n.s.	n.s.	n.s.	n.s.	n.s.	0.7421
5	-1.18323	n.s.	n.s.	n.s.	n.s.	n.s.	0.88488
6	-1.04686	-0.00645	n.s.	n.s.	n.s.	n.s.	0.99332
7	-0.57081	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
8	-1.1725	n.s.	n.s.	n.s.	n.s.	n.s.	1.05233
9	-1.38982	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
10	-1.12776	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
11	-1.24097	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
12	-0.91632	-0.00493	n.s.	n.s.	n.s.	n.s.	1.01626
13	-1.21374	-0.00508	n.s.	n.s.	0.066558	n.s.	1.34313
14	-1.02541	-0.00631	n.s.	0.04579	n.s.	n.s.	0.81935
15	-0.97745	-0.0066	n.s.	n.s.	n.s.	n.s.	1.5062
16	-1.08256	n.s.	n.s.	n.s.	n.s.	n.s.	0.71314
17	-0.99552	-0.00481	n.s.	n.s.	n.s.	n.s.	0.94613
18	-1.21922	n.s.	n.s.	n.s.	n.s.	n.s.	0.74115
19	-1.13076	-0.00523	n.s.	n.s.	n.s.	n.s.	1.14734
20	-1.20248	n.s.	n.s.	0.04831	n.s.	n.s.	n.s.

Analizzando i risultati ottenuti è possibile fare le seguenti considerazioni:

- le variabili *anno di posa*, *pressione massima*, *velocità media* e *flusso veicolare* non risultano significative almeno 18 volte su 20;
- in tutte le estrazioni la variabile *materiale* risulta rilevante per il modello;
- la variabile *diametro* non risulta significativa in 10 estrazioni, mentre *Tram Route* non è una variabile rilevante per il modello in 5 estrazioni su 20;
- in tutte le estrazioni il coefficiente β della variabile *tram route* ha segno positivo: la probabilità di rottura aumenta se la variabile ha valore pari ad uno, indice della presenza del tram;
- in tutte le estrazioni i coefficienti β delle variabili *materiale* e *diametro* hanno segno negativo: la probabilità di rottura diminuisce all'aumentare dei valori che le variabili assumono.

Infine, si riportano i grafici a bolle ottenuti dall'applicazione dei modelli stimati sul 20% delle estrazioni, ossia i set di test. Ogni grafico rappresenta la probabilità di rottura stimata in funzione della classe di guasto ($flag=1$; $flag=0$) e le dimensioni delle bolle sono direttamente proporzionali alle occorrenze.

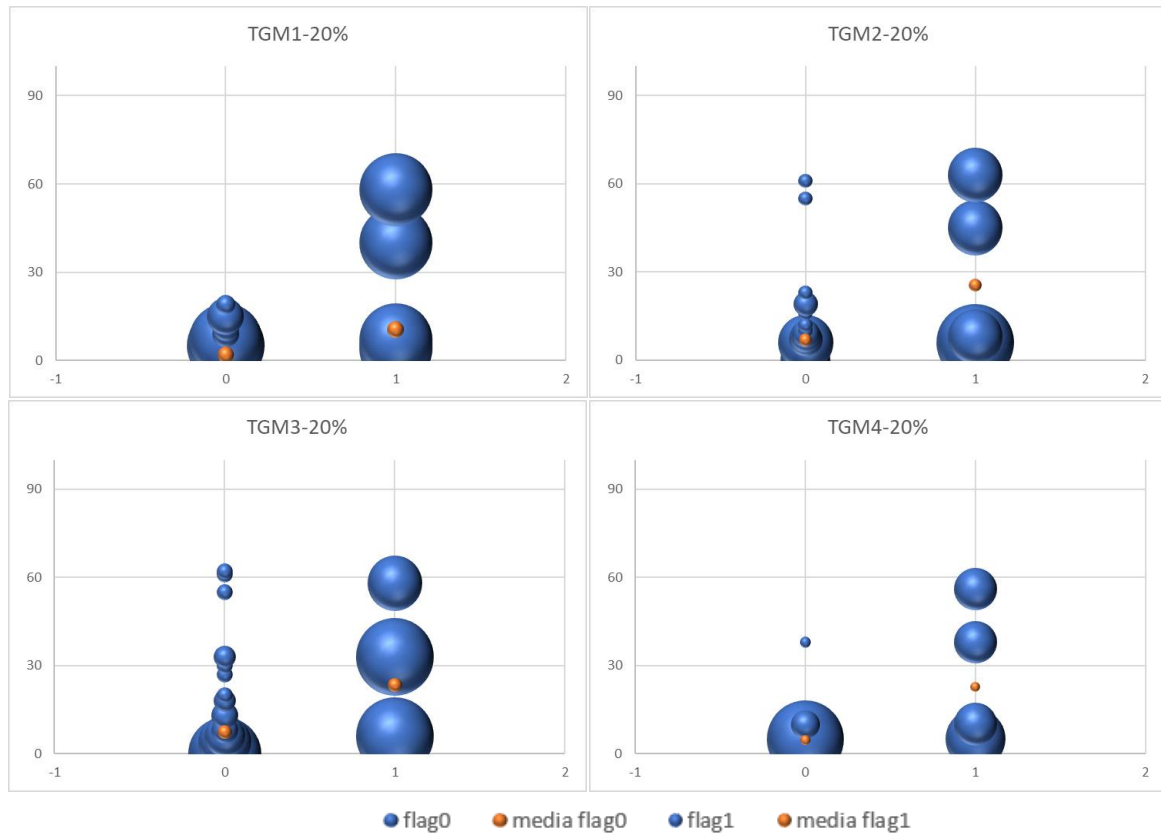


Figura 6-24 Grafici a bolle ottenuti dalla fase di test del modello

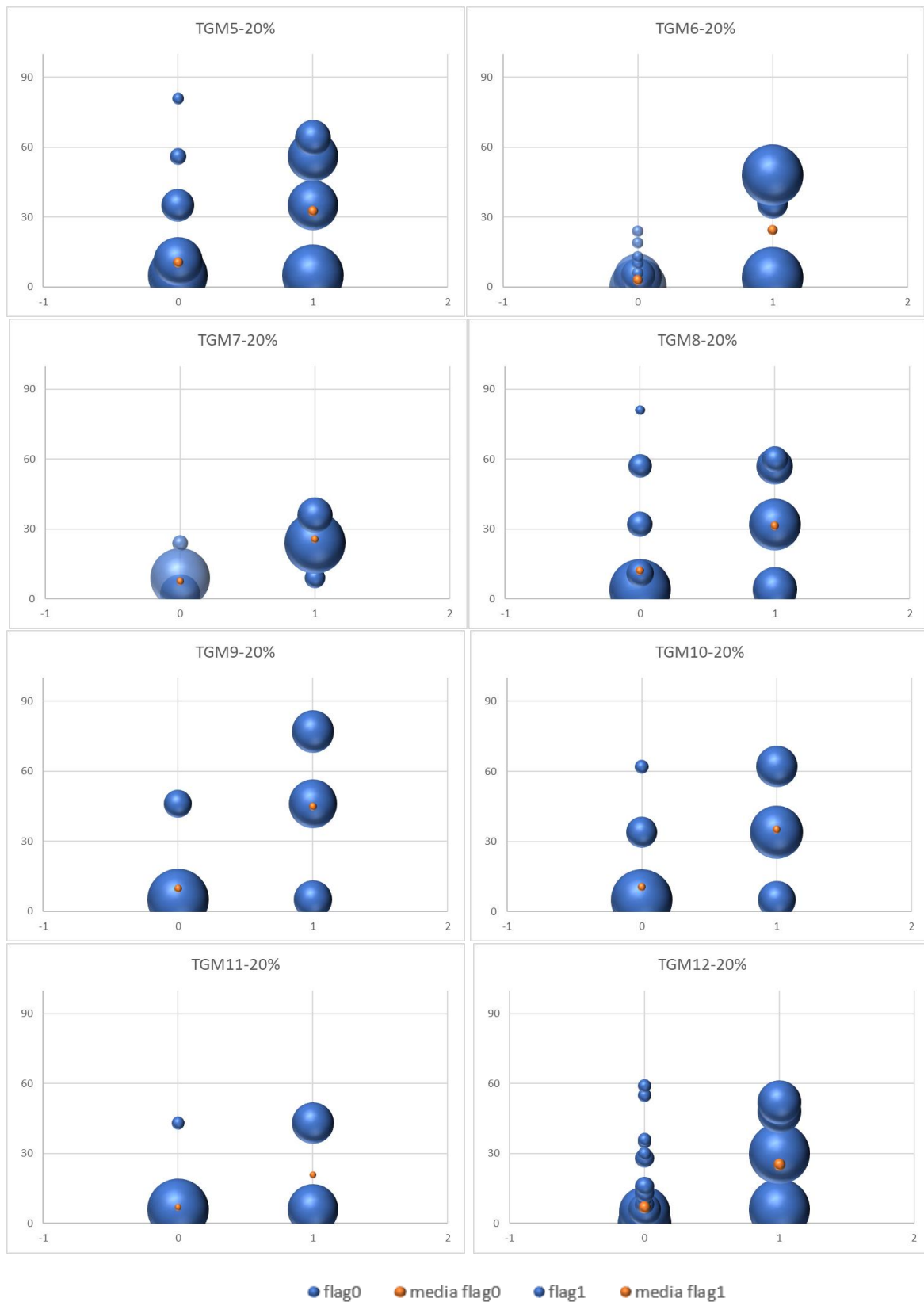


Figura 6-25 Grafici a bolle ottenuti dalla fase di test del modello

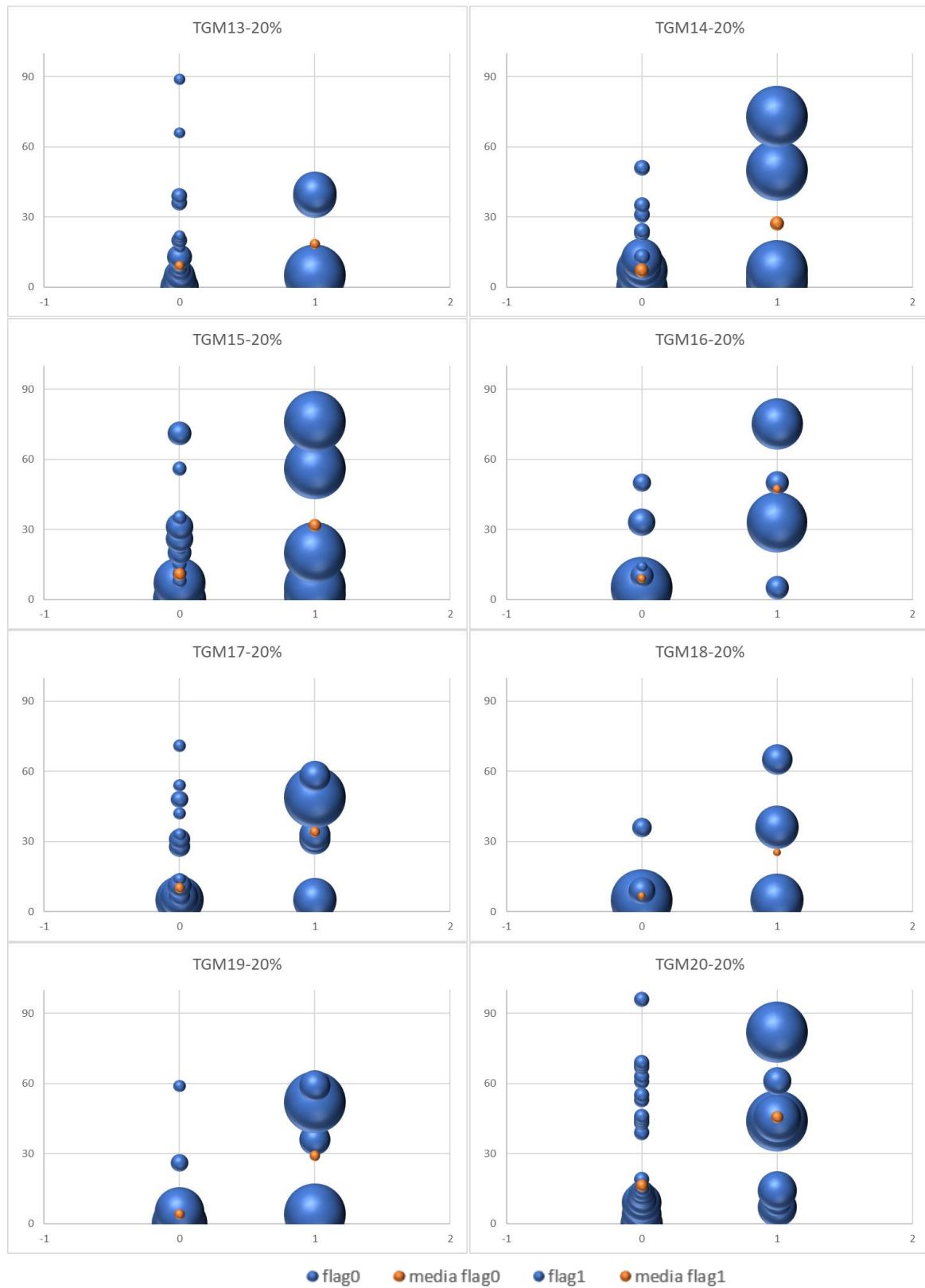


Figura 6-26 Grafici a bolle ottenuti dalla fase di test del modello

Si osserva che il valore medio della probabilità di rottura stimata dal modello è sempre minore per le condotte che hanno flag pari a 0.

A conferma di ciò si riportano su un unico grafico a bolle tutti i risultati ottenuti applicando la relazione per il calcolo della probabilità sugli elementi delle estrazioni al 20%.

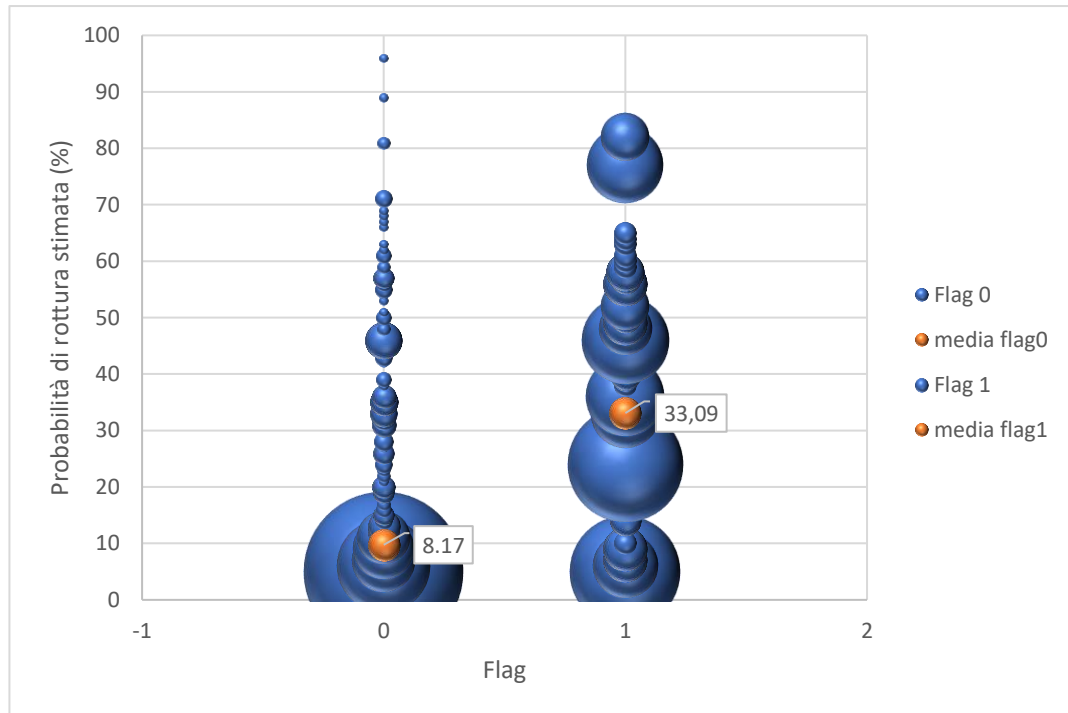


Figura 6-27 Probabilità di rottura stimata relativa ai 20 set di test

I risultati della validazione del modello sui venti set di dati consentono di affermare che la previsione delle rotture può essere spiegata dal modello di regressione logistica, tenuto conto delle variabili significative.

Il modello soddisfa, quindi, la verifica di capacità previsiva e poiché è valido su più estrazioni, risulta estensibile alla tabella *RegressioneTGM*, ovvero gode di possibili generalizzazioni.

6.3 IL MODELLO GENERALE APPLICATO ALLA TABELLA REGRESSIONETGM

Gli esiti precedentemente ottenuti autorizzano l'applicazione del modello di regressione logistica sull'intera tabella *RegressioneTGM*. Valgono le stesse considerazioni fatte precedentemente riguardo l'esclusione della variabile 'variazione di carico' e riguardo l'assegnazione di un numero intero per codificare il campo 'materiale'; le variabili del modello sono:

- variabile risposta dipendente Y : flag (1=rottura; 0=non rottura);
- variabile indipendente x_1 : numero materiale;
- variabile indipendente x_2 : diametro;
- variabile indipendente x_3 : anno di posa;
- variabile indipendente x_4 : pressione massima di esercizio;
- variabile indipendente x_5 : velocità media;
- variabile indipendente x_6 : flusso veicolare;
- variabile indipendente x_7 : Tram Route;

Viene riportato l'output fornito dal software *Gretl*, scegliendo in prima approssimazione di introdurre tutte le potenziali variabili.

```
Modello 1: Logit multinomiale, usando le osservazioni 1-2562
Variabile dipendente: flag
Errori standard basati sull'Hessiana
```

	coefficiente	errore std.	z	p-value	

flag = 1					
const	-8,63355	8,14826	-1,060	0,2893	
N_materiale	-1,26815	0,193719	-6,546	5,90e-011	***
diámetro	-0,00565504	0,00206938	-2,733	0,0063	***
anno_posa	0,00467903	0,00433925	1,078	0,2809	
pMax	0,000187337	0,00768819	0,02437	0,9806	
velocità	0,00130266	0,0278722	0,04674	0,9627	
flusso	0,000338198	0,000625962	0,5403	0,5890	
TramRoute	0,980479	0,309382	3,169	0,0015	***

Media var. dipendente	0,024590	SQM var. dipendente	0,154903		
Log-verosimiglianza	-223,6056	Criterio di Akaike	463,2113		
Criterio di Schwarz	509,9996	Hannan-Quinn	480,1765		
Note: SQM = scarto quadratico medio; E.S. = errore standard					
Numero dei casi 'previsti correttamente' = 2499 (97,5%)					
Test del rapporto di verosimiglianza: Chi-quadro(7) = 144,108 [0,0000]					

Figura 6-28 Stime dei parametri del modello di regressione logistica applicato alla tabella *RegressioneTGM*

In Figura 6-28 l'indicatore visivo e i valori della colonna '*p-value*', maggiori in alcuni casi del livello di significatività imposto, pari a 0.05, rivelano che alcune delle variabili introdotte non rappresentano fattori determinanti della dicotomica, ossia non risultano significative per il modello. Si procede, quindi, omettendo progressivamente tali variabili.

```

Eliminazione sequenziale usando alfa (a due code) = 0,05

Scartata pMax          (p-value 0,981)
Scartata velocitA      (p-value 0,953)
Scartata flusso        (p-value 0,507)
Scartata anno_posa     (p-value 0,253)

Test sul modello 1

Ipotesi nulla: i parametri della regressione valgono zero per le variabili
anno_posa, pMax, velocitA, flusso
Statistica test: F(4, 2554) = 0,432181, p-value 0,785477
L'omissione delle variabili ha migliorato 3 dei 3 criteri di informazione.

Modello 2: Logit multinomiale, usando le osservazioni 1-2562
Variabile dipendente: flag
Errori standard basati sull'Hessiana

-----
                coefficiente    errore std.    z        p-value
-----
flag = 1
const          0,270372        0,445500    0,6069    0,5439
N_materiale    -1,11579         0,150812   -7,399    1,38e-013 ***
diametro       -0,00547254         0,00204527 -2,676    0,0075 ***
TramRoute      0,948453          0,273208    3,472    0,0005 ***

Media var. dipendente  0,024590    SQM var. dipendente  0,154903
Log-verosimiglianza   -224,4571    Criterio di Akaike   456,9142
Criterio di Schwarz    480,3083    Hannan-Quinn         465,3968
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 2499 (97,5%)
Test del rapporto di verosimiglianza: Chi-quadro(3) = 142,405 [0,0000]

```

Figura 6-29 Stime dei parametri del modello applicato alla tabella *RegressioneTGM* con variabili significative

Tra i requisiti fondamentali di un buon modello vi è la *parsimonia*, che prevede un numero esiguo di variabili dipendenti, al fine di rendere il modello stabile e facilmente interpretabile. L'insieme delle variabili esplicative è composto da: '*materiale*', '*diametro*' e '*tram route*', vengono scartate, in linea con i risultati ottenuti precedentemente, le restanti variabili.

Risulta complesso definire perché il modello su base statistica elimini le variabili di traffico '*velocità media*' e '*flusso veicolare*', ma definisca significativa '*tram route*'. Una possibile motivazione potrebbe essere ricercata nella capacità dei materiali di assorbire le vibrazioni. Le condotte presenti nella tabella *RegressioneTGM* sono per il 63% costituite da ghisa sferoidale, che è un materiale stabile in condizioni di traffico intenso.

Una spiegazione potrebbe darla anche l'esistenza di differenze tra le vibrazioni prodotte dal tram e dai veicoli su gomma. In linea di principio un veicolo con carico bilanciato, che procede su una strada priva di imperfezioni, non produce vibrazioni. Nella realtà il manto

stradale è soggetto ad usura ed il transito è discontinuo. Le vibrazioni trasmesse al terreno dipendono quindi molto dallo strato di asfalto. In città le vibrazioni prodotte dal transito di tram possono essere paragonate a quelle di un treno, che procede a velocità molto ridotta e che subisce frequentemente accelerazioni e rallentamenti. Lo stress generato dal tram è individuato da numerosi fattori, come la disomogeneità del terreno e la variabilità di pavimentazione sotto l'armamento tranviario, ma, a differenza del veicolo su gomma, anche dalle forze laterali generate dai vincoli che la rotaia impone alla ruota.

Definite le variabili significative, il modello di regressione logistica assume la forma:

$$\text{logit}[P(Y = 1)] = +0.2703 - 1.1157x_1 - 0.0054x_2 + 0.9484x_7$$

Il segno algebrico dei coefficienti fornisce precisazioni, fondamentali e importanti per comprendere la funzionalità del modello:

- il segno negativo di β_1 indica che all'aumentare del numero assegnato a ciascun materiale, la probabilità di rottura diminuisce;
- il segno negativo di β_2 indica che all'aumentare del diametro della condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_7 indica che la presenza del tram fa sì che la probabilità di rottura aumenti.

Noti i coefficienti di regressione, la probabilità di rottura stimata è calcolata per ogni condotta presente in tabella come:

$$\hat{\pi}(x) = \frac{e^{0.2703 - 1.1157x_1 - 0.0054x_2 + 0.9484x_7}}{1 + e^{0.2703 - 1.1157x_1 - 0.0054x_2 + 0.9484x_7}}$$

Per verificare la capacità predittiva del modello, i valori di probabilità di rottura stimati sono stati riportati su un grafico a bolle, dove le dimensioni di ogni bolla sono proporzionali all'occorrenza.

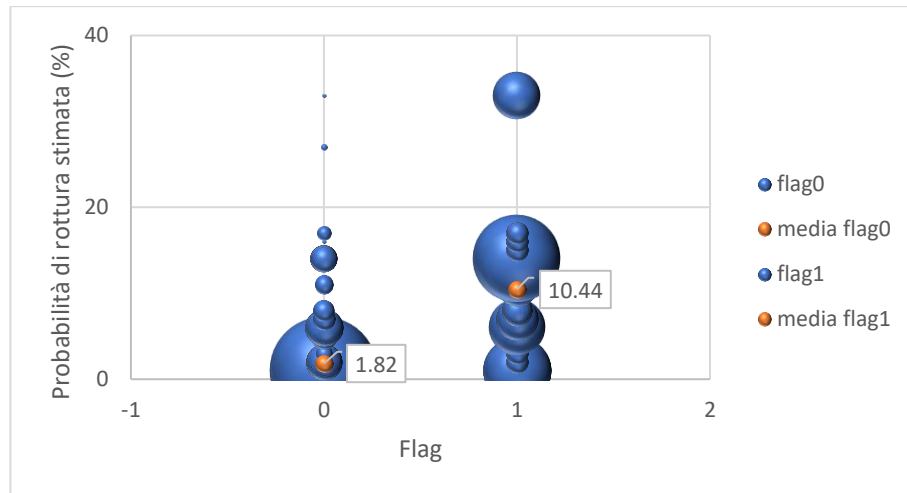


Figura 6-30 Confronto tra la probabilità di rottura stimata e il dato reale delle condotte della tabella *RegressioneTGM*

La verifica del modello è soddisfatta, in quanto le stime di probabilità ottenute sono il più possibile vicine ai valori realmente osservati, come confermano i valori medi di probabilità, e si parla quindi di *bontà dell'adattamento*.

6.4 CONCLUSIONI

In questo capitolo è stato applicato il modello di regressione logistica ai dati a disposizione. Il modello è stato inizialmente definito su 20 set di dati, estratti dalla tabella *RegressioneTGM* in modo che fossero caratterizzati da un valore percentuale di rottura pari a quello della tabella *ReteTorino*. Ogni set di dati è costituito da:

- 63 condotte guaste;
- 357 condotte non guaste, estratte casualmente dalla tabella *RegressioneTGM*.

Tenuto conto della correlazione fra le variabili, sono stati selezionati i regressori e dopo aver applicato il modello ad ognuna delle 20 estrazioni, sono stati letti i risultati e interpretate le stime dei singoli coefficienti.

Per ben capire le prestazioni del modello, i 20 set di dati sono stati divisi in un set di allenamento, sul quale viene tarato il modello, pari all'80% degli elementi, e un set di test, sul quale avviene la previsione del modello, pari al restante 20% degli elementi di ogni estrazione. Ogni nuovo campione ha una percentuale di rottura diversa, ottenuta dalla selezione casuale delle condotte.

La stima dei coefficienti di regressione avviene applicando il metodo *backward*, che elimina le variabili non significative, pur prendendo inizialmente in considerazione tutti i regressori.

Poiché il valore medio della probabilità di rottura stimata dal modello è, nei venti set al 20%, maggiore per le condotte che hanno flag pari a 1, e quindi soddisfa la bontà di

adattamento, si estende il modello alla tabella *RegressioneTGM*. Dopo aver esaminato i risultati forniti da *Gretl*, si riscontra che le sole variabili significative sono:

- materiale;
- diametro;
- tram route.

Mentre il valore di probabilità di rottura delle condotte con flag pari a 0, è, anche in questo caso, inferiore se confrontata con la probabilità stimata per le condotte che hanno presentato un guasto. Il modello ben si presta a rappresentare i dati.

7 ESTENSIONE DEL MODELLO E DEFINIZIONE DEL VALORE DI SOGLIA

L'analisi condotta ha ristretto il campo a 3 variabili indipendenti e ha consentito la stima della probabilità di rottura delle condotte appartenenti alla tabella *RegressioneTGM*.

La descrizione della variabile Y , ottenuta applicando il modello di regressione logistica, studiata in un contesto di 2562 condotte, è estendibile ad altri campioni con margini d'errore ragionevoli. Integrando la tabella *RegressioneCondotte* con l'informazione della presenza del tram, si ottiene un set di dati costituito da 5740 condotte sul quale stimare i parametri di regressione.

7.1 ANALISI DELLA VARIABILE 'TRAM ROUTE'

La tabella *RegressioneCondotte* è costituita da:

- 5740 condotte;
- 5653 condotte non rotte ($flag=0$);
- 87 condotte rotte ($flag=1$);
- 106 rotture;
- percentuale di rottura pari al 1.85%.

La realizzazione di questa ha previsto l'estrazione di sole quelle condotte che possiedono tutti i valori di: *anno di posa*, *diametro*, *materiale*, *pressione massima* e *variazione di carico* dalla tabella *ReteTorino*. Poiché il modello di regressione definisce significativa la variabile 'tram route', utilizzando i dati sul servizio *GTT (Gruppo Torinese Trasporti)*, forniti dalla società stessa, è stato aggiunto un campo alla tabella *RegressioneCondotte* per indicare la presenza di binari dedicati al passaggio dei tram e quindi se la tubazione sottostante fosse soggetta a carichi dinamici generati dal sistema tranviario.

Il campo aggiunto, denominato similmente alla tabella *RegressioneTGM*, assume valore pari a:

- 1 in presenza della linea tranviaria;
- 0 in assenza della linea tranviaria.

Al fine di studiare l'andamento della nuova variabile nella tabella *RegressioneCondotte* e calcolare il tasso di fallanza, le condotte:

- sono state suddivise in classi in funzione della presenza/assenza del tram;
- per ogni classe è stata calcolata la lunghezza complessiva delle condotte e il numero di rotture che le appartengono (condotte con flag pari a 1);
- è stata applicata la relazione (1), ricavando il tasso di fallanza per ogni classe.

Si riportano in Tabella 7-1 i risultati ottenuti:

Tabella 7-1 Calcolo del tasso di fallanza delle condotte soggette (1) o non soggette (0) alle vibrazioni indotte dal tram

Tram Route	Lunghezza [km]	%	N° di guasti	% guasti	Tasso di fallanza [NR/(km*anno)]
0	254.397	89.047	74.000	69.811	0.291
1	31.293	10.953	32.000	30.189	1.023
Totale	285.69	100	106	100	

In Figura 7-1 si riportano, per via grafica, tali risultati per permettere una valutazione più immediata del tasso di rottura.

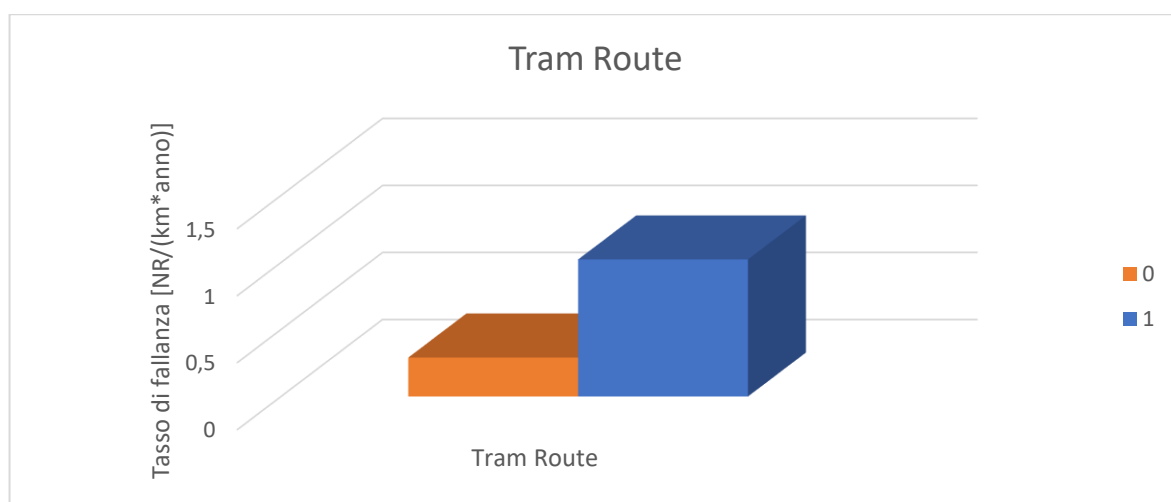


Figura 7-1 Tasso di fallanza delle classi di presenza/assenza tram nella tabella *RegressioneCondotte*

L'89% in lunghezza delle condotte che popolano la tabella *RegressioneCondotte* non è soggetto a carichi dovuti al passaggio del tram; ciò nonostante in questa classe si sono verificati 74 guasti, ma il valore del tasso di fallanza più alto è individuato nella classe 1.

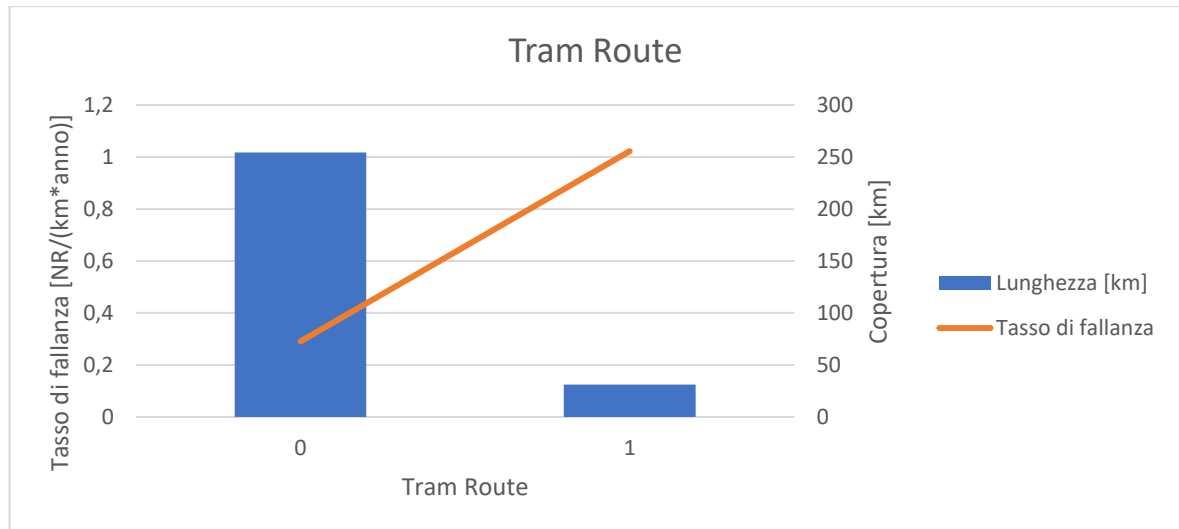


Figura 7-2 Tasso di rottura osservato e copertura delle classi di condotte soggette (1) o non (0) al passaggio del tram

7.2 APPLICAZIONE DEL MODELLO

L'applicazione del modello di regressione prevede che siano definite le variabili sulle quali avverrà la stima dei coefficienti di regressione; queste sono:

- variabile risposta dipendente Y : flag (1=rottura; 0=non rottura);
- variabile indipendente x_1 : numero materiale;
- variabile indipendente x_2 : diametro;
- variabile indipendente x_3 : anno di posa;
- variabile indipendente x_4 : carico massimo;
- variabile indipendente x_5 : tram route.

Viene riportato l'output del software *Gretl*, in cui si è già applicato il metodo *backward*. Il criterio utilizzato per definire le variabili significative, ha scartato 'carico massimo' e 'anno di posa' perché caratterizzati da un valore del *p-value* maggiore del livello di significatività imposto. Confermando i risultati ottenuti nel Capitolo 6, l'insieme delle variabili significative è composto da: 'materiale', 'diametro' e 'tram route', di cui sono ora noti i coefficienti di regressione.

Il modello di regressione logistica assume la forma:

$$\text{logit}[P(Y = 1)] = 0.2280 - 1.0416x_1 - 0.0057x_2 + 0.7943x_7$$

```

Eliminazione sequenziale usando alfa (a due code) = 0,05

Scartata anno_posa      (p-value 0,861)
Scartata pMax           (p-value 0,437)

Test sul modello 5

Ipotesi nulla: i parametri della regressione valgono zero per le variabili
anno_posa, pMax
Statistica test: F(2, 5753) = 0,31741, p-value 0,728045
L'omissione delle variabili ha migliorato 3 dei 3 criteri di informazione.

Modello 6: Logit multinomiale, usando le osservazioni 1-5759
Variabile dipendente: flag
Errori standard basati sull'Hessiana

      coefficiente    errore std.      z      p-value
-----
flag = 1
const      0,228099      0,347404      0,6566    0,5114
N_materiale -1,04167      0,101088     -10,30    6,72e-025 ***
diametro   -0,00570606      0,00179493    -3,179    0,0015    ***
TramRoute    0,794354      0,238683      3,328    0,0009    ***

Media var. dipendente  0,018406    SQM var. dipendente  0,134426
Log-verosimiglianza   -432,7093    Criterio di Akaike   873,4185
Criterio di Schwarz    900,0526    Hannan-Quinn         882,6869
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 5652 (98,1%)
Test del rapporto di verosimiglianza: Chi-quadro(3) = 191,575 [0,0000]

```

Figura 96 Stime dei parametri del modello applicato alla tabella *RegressioneCondotte* potenziata dal campo 'tram route' con variabili significative

Il segno algebrico dei coefficienti fornisce precisazioni, fondamentali e importanti per comprendere la funzionalità del modello:

- il segno negativo di β_1 indica che all'aumentare del numero assegnato a ciascun materiale, la probabilità di rottura diminuisce;
- il segno negativo di β_2 indica che all'aumentare del diametro della condotta, la probabilità di rottura diminuisce;
- il segno positivo di β_7 indica che la presenza del tram fa sì che la probabilità di rottura aumenti.

Il software fornisce il valore della statistica *Chi-quadro* alla voce 'Test del rapporto verosimiglianze'. Il valore restituito dal modello, con 3 gradi di libertà, riportati fra parentesi tonde, soddisfa la relazione:

$$G = 191.575 > \chi_{\frac{\alpha}{2}}^2 = 12.38$$

Inoltre il valore del *p-value*, riportato fra parentesi quadre, è minore di 0.05, conseguentemente l'ipotesi nulla può essere rifiutata.

Noti i coefficienti di regressione, la probabilità di rottura stimata è calcolata per ogni condotta presente in tabella come:

$$\hat{\pi}(x) = \frac{e^{0.2280 - 1.0416x_1 - 0.0057x_2 + 0.7943x_7}}{1 + e^{0.2280 - 1.0416x_1 - 0.0057x_2 + 0.7943x_7}}$$

Dato che i valori stimati sono probabilità, le condotte possono essere classificate in funzione del flag secondo la probabilità predetta.

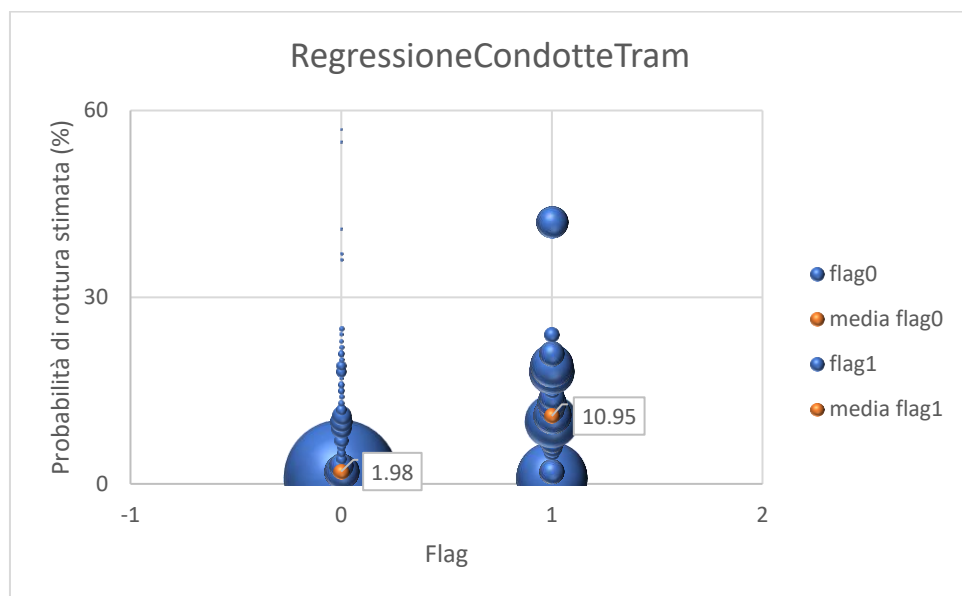


Figura 7-4 Confronto tra la probabilità di rottura stimata e il dato reale delle condotte della tabella *RegressioneCondotte* potenziata dal campo 'tram route'

Si osserva che il valore medio della probabilità di rottura stimata dal modello è minore per le condotte che hanno flag pari a 0. I risultati ottenuti sulla tabella *RegressioneCondotte*, in cui è stata inclusa l'informazione sulla presenza del tram, consentono di affermare che la previsione delle rotture può essere spiegata dal modello di regressione logistica, tenuto conto delle variabili significative.

7.3 DEFINIZIONE DEL VALORE DI SOGLIA

Una volta applicata la relazione che consente di valutare la probabilità di rottura stimata di singole condotte, si dispone di un metodo di classificazione, che ha l'obiettivo di identificare ed associare una priorità di intervento per la manutenzione delle stesse.

In una situazione ideale ci si aspetta che i valori ottenuti dal modello siano in grado di discriminare perfettamente le due classi (flag=1; flag=0), che costituiscono il set di dati, fissato il cut-off, ossia un valore soglia del test. In realtà ciò che avviene è una parziale sovrapposizione delle due classi, infatti il modello identifica come positive alcune condotte non guaste (Falsi Positivi) e come negative alcune condotte che hanno nella realtà presentato rottura (Falsi Negativi). La scelta del valore di cut-off influenza il rischio di valutare un numero maggiore o minore di FP o di FN.

In proposito, definito un range di variazione della probabilità di soglia compreso tra il 10% e il 30%, sono state calcolate le occorrenze delle condotte risultanti falsi negativi e falsi positivi ed è stata calcolata la loro somma per ogni valore dell'intervallo:

- sono falsi positivi (FP) le condotte con flag pari a 0 e probabilità di rottura stimata maggiore o uguale ad un valore definito, appartenente al range. Queste non hanno presentato rottura, ma sono considerate suscettibili dal modello di regressione;
- sono falsi negativi (FN) le condotte con flag pari a 1 e probabilità di rottura stimata minore o uguale ad un valore definito, appartenente al range. Queste hanno realmente presentato rottura, ma non sono considerate suscettibili dal modello.

Tabella 7-2 Numero di falsi positivi e falsi negativi al variare del cut off

Cut-off (%)	FP	FN	Totale
10	368	57	425
11	246	64	310
12	157	67	224
13	128	71	199
14	120	74	194
15	114	75	189
16	103	77	180
17	93	77	170
18	90	87	177
19	63	96	159
20	41	97	138
21	36	100	136
22	25	100	125
23	20	100	120
24	16	101	117
25	13	101	114
26	7	101	108
27	7	101	108
28	7	101	108
29	7	101	108
30	7	101	108

Il criterio adottato per la definizione del cut-off ottimale prevede di considerare il valore di probabilità che minimizza la sommatoria tra falsi positivi e falsi negativi, ossia l'ascissa del primo minimo locale della funzione rappresentata in Figura 7-5.

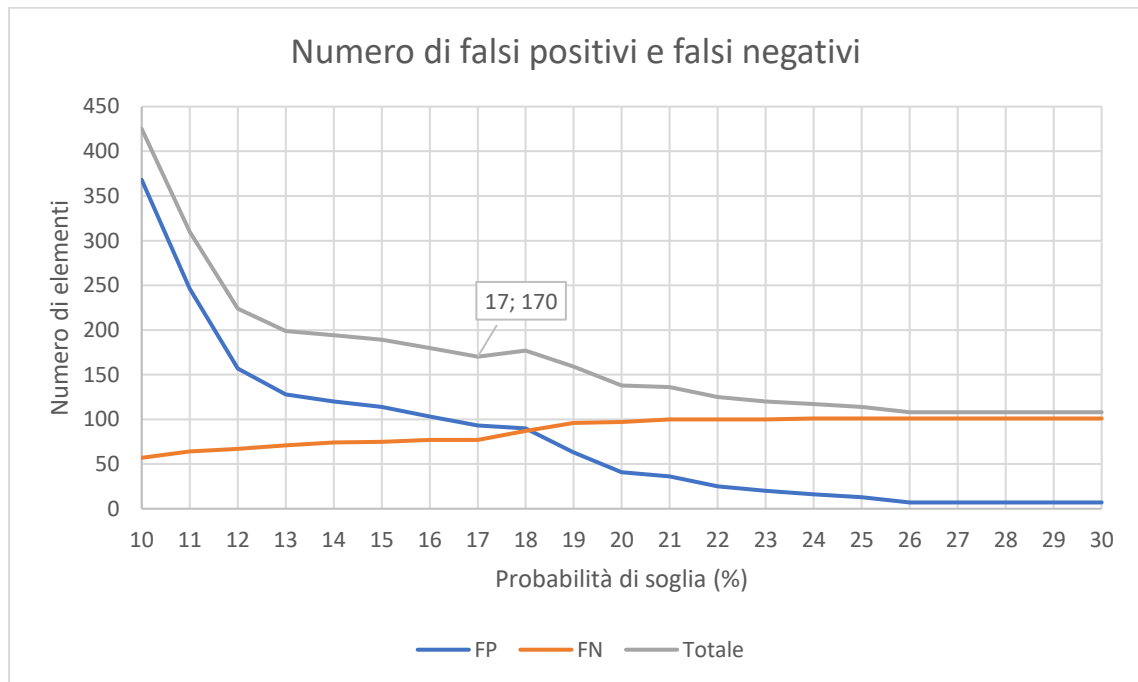


Figura 7-5 Curva dei FP, FN e della loro somma al variare della probabilità di soglia

Il valore di soglia ottimale, superato il quale una condotta può definirsi suscettibile di rottura, adottando questo criterio, è pari al 17%, a cui corrisponde un totale di falsi positivi e falsi negativi pari a 170.

È stato quindi possibile categorizzare la rete di distribuzione di Torino e visualizzarla su mappa, tramite l'applicazione *QGIS Desktop*, in funzione della probabilità di rottura stimata (Figura 7-6).

A tal proposito sono state definite quattro classi a cui è stato assegnato un colore: nella prima vi sono le condotte non suscettibili di rottura, nelle successive vi sono condotte con probabilità stimata via via crescente.

7.4 CONCLUSIONI

Considerata la bontà di adattamento e la capacità previsiva del modello applicato alla *TabellaTGM*, si è ritenuto opportuno estendere l'analisi ad un set di dati più numeroso: la tabella *RegressioneCondotte*. Il primo passo è stato l'annessione alla tabella del campo 'tram route', variabile giudicata significativa dalle applicazioni precedenti del modello. È stato valutato l'andamento della variabile e, a tal proposito, l'analisi del tasso di rottura consente di affermare che la classe di condotte soggette al passaggio del tram, presenta un valore più alto, mentre la classe delle condotte non soggette al passaggio risulta la più estesa in chilometri.

L'applicazione del modello di regressione logistica conferma i risultati ottenuti in precedenza; le variabili significative sono:

- materiale;
- diametro;
- tram route.

Il calcolo della probabilità di rottura stimata, utilizzando i coefficienti di regressione, mostra come la maggior parte delle condotte non guaste presenti probabilità inferiore rispetto a quella presentata dalle condotte realmente guaste.

Per concludere, è stata valutata la probabilità di soglia, che costituisce, in questo contesto, un indicatore di rottura. Il criterio scelto per l'individuazione del valore si basa sul calcolo del numero di condotte definibili falsi positivi e falsi negativi, e sulla ricerca del primo valore che minimizza la loro somma. Il valore ottimale di cut-off identificato è pari al 17%. Il criterio proposto è uno dei tanti attuabili, in quanto la scelta deve necessariamente essere affiancata da un'analisi economica costi/benefici, che consenta di valutare la complessità del problema.

CONCLUSIONI

Questa tesi presenta l'applicazione di un modello di regressione logistica alla rete idrica di Torino con lo scopo di evidenziare fattori significativi nella previsione della probabilità di rottura delle condotte.

La pianificazione efficace della manutenzione e riabilitazione del sistema di distribuzione dell'acqua richiede un accurato studio del deterioramento strutturale della rete idrica. Un sistema di distribuzione è composto da centinaia e persino migliaia di condotte interrato che lavorano in pressione; il monitoraggio dello stato strutturale è spesso proibitivamente costoso, a causa delle condizioni operative, che lo rendono di non facile esecuzione. La misura del deterioramento strutturale di un sistema può essere individuata utilizzando la stima delle probabilità delle rotture nel tempo, che costituisce un'alternativa efficace ed economica.

Un modello in grado di quantificare la probabilità di rottura di una condotta, deve considerare caratteristiche geometriche, proprietà operative, informazioni storiche sui guasti e tassi di rottura, ma anche fattori dipendenti dall'ambiente in cui la condotta è inserita. L'influenza di ciascuno di questi fattori sul tasso di rottura può essere quantificata per supportare il gestore nella pianificazione delle strategie di manutenzione, atte a limitare i futuri tassi di rottura.

A tal proposito la prima parte del lavoro di tesi è stata dedicata alla descrizione della rete idrica di Torino e all'organizzazione delle informazioni sulle condotte e sui guasti, avvenuti nel decennio 2006-2016, raccolte negli anni dal personale della *SMAT* in un unico database (la piattaforma Maximo).

Si è preferito integrare le condizioni strutturali e idrauliche delle condotte con condizioni esterne di traffico, le cui informazioni sono state fornite dalle società *5T* e *GTT*. La scelta in questione definisce in modo definitivo i parametri che hanno rappresentato le variabili predittive, di natura qualitativa e quantitativa, dell'analisi statistica, quali: materiale, diametro, anno di posa, pressione massima, variazione di pressione, velocità media e flusso medio veicolare, presenza/assenza tram. La variabile dipendente dicotomica è rappresentata dalla presenza/assenza di guasti, in quanto in questo contesto applicativo può assumere solo due valori (rottura=1; non rottura=0).

Il modello scelto per condurre l'analisi non prevede che vi sia carenza di informazioni in ogni campo del dataset; per questo motivo i dati di partenza sono stati interessati da importanti operazioni di filtraggio, atti dapprima ad eliminare le condotte prive di almeno un'informazione strutturale ed operativa, in seguito ad eliminare le condotte per cui non vi era perfetta sovrapposizione con le informazioni di traffico degli archi stradali a disposizione; tali operazioni coincidono con le fasi realizzative della tabella *RegressioneTGM*.

L'analisi della tabella, sulla quale è possibile applicare il modello, ha previsto il calcolo del tasso di fallanza delle condotte suddivise in classi in funzione delle variabili predittive. I risultati consentono di affermare che:

- i materiali più vulnerabili sono l'eternit e la ghisa grigia;
- la ghisa grigia è il materiale più presente in rete;
- il tasso di fallanza diminuisce all'aumentare del diametro. I diametri in cui sono state registrate il numero maggiore di rotture sono quelli compresi tra 100 mm e 150 mm;
- le condotte posate nel ventennio 1930-1950 sono le più vulnerabili;
- gran parte delle condotte sono ubicate in corrispondenza di strade percorribili a media-bassa velocità e caratterizzate da flussi veicolari ridotti;
- le condotte posate in strade soggette al passaggio del tram hanno tasso di fallanza maggiore.

Il calcolo dei coefficienti di correlazione conduce a scartare una tra le variabili '*pressione massima*' e '*variazione di carico*', poiché fortemente correlate e non utilizzabili contemporaneamente nello sviluppo della regressione logistica.

È stata dunque svolta l'analisi di previsione e per dimostrare l'efficacia del modello di regressione, la tesi include l'analisi di 20 estrazioni, caratterizzate da un valore percentuale di rottura pari a quello dell'intera rete di Torino. Dalla stima e l'interpretazione dei coefficienti di regressione si evince che:

- le variabili *pressione massima*, *velocità media* e *flusso veicolare* risultano sempre poco significative;
- in tutte le estrazioni la variabile *materiale* risulta rilevante per il modello;
- le variabili *diametro* e *tram route* non sono una variabile significativa per un quarto delle estrazioni;
- in una sola estrazione la variabile *anno di posa* risulta significativa;
- in tutte le estrazioni il coefficiente β della variabile *tram route* ha segno positivo; la probabilità di rottura aumenta se la variabile ha valore pari ad uno, indice della presenza del tram;
- in tutte le estrazioni i coefficienti β delle variabili *materiale* e *diametro* hanno segno negativo; la probabilità di rottura diminuisce all'aumentare dei valori che le variabili assumono.

Per ben capire le prestazioni del modello, i 20 set di dati sono stati divisi con rapporto 80:20, sulla prima parte viene tarato il modello e sulla seconda avviene il test e la previsione. Ogni nuovo campione ha una percentuale di rottura diversa, ottenuta dalla selezione casuale delle condotte. La stima dei coefficienti di regressione avviene applicando il metodo *backward*, che elimina le variabili non significative, pur prendendo inizialmente in considerazione tutte le predittive. Poiché il modello soddisfa la bontà di adattamento è stato esteso alla tabella *RegressioneTGM*. I risultati forniti dal software *Gretl*, individuano le seguenti variabili significative:

- materiale;
- diametro;
- tram route.

Mentre le stime di probabilità ottenute sono il più possibile vicine ai valori realmente osservati, come confermato dai valori medi di probabilità.

Vista la significatività della variabile '*tram route*' e la capacità previsiva del modello, si è ritenuto opportuno estendere l'analisi ad un dataset più numeroso perché privo delle informazioni di traffico, quali velocità media e flusso medio veicolare, che durante le operazioni di filtraggio avevano causato l'eliminazione di molte condotte.

L'applicazione del modello di regressione logistica ha confermato i risultati ottenuti in precedenza in termini di significatività delle variabili. Il modello associa, inoltre, a condotte realmente rotte un valore di probabilità media stimata di rottura maggiore. Infine, è stato valutato il valore di cut off, attraverso un criterio basato sul calcolo del numero di condotte definibili falsi positivi e falsi negativi, e sulla ricerca del primo valore che minimizza la loro somma. Il valore soglia ottimale individuato è pari al 17%. Il criterio proposto non è l'unico applicabile, e la sua scelta ha ripercussioni economiche, sociali ed ambientali e non può prescindere dalle politiche attuative del gestore. I risultati ottenuti sono stati riportati su mappa per consentire la visualizzazione delle zone della città maggiormente esposte al rischio rottura.

Tuttavia, è importante tenere presente che l'analisi si è concentrata su un numero ridotto di guasti, infatti la carenza di dati di rottura e di dati di traffico, non estesi a tutta la rete stradale, hanno costituito un limite per una più profonda comprensione del processo.

Con più dati, completi di ogni informazione nei campi, la fiducia nelle previsioni generate dal modello di regressione proposto sarà maggiore.

APPENDICE A

Si riportano nel dettaglio i passi eseguiti, nella fase di fitting e di testing, per l'applicazione del modello. I valori della 'Estrazione_TGM1-80%' sono stati presentati in §6.2.

Estrazione TGM2-80%

Si riporta l'output fornito dal software *Gretl* prendendo in considerazione tutte le variabili esplicative.

```
Modello 1: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

-----
               coefficiente      errore std.      z      p-value
-----
flag = 1
const          -30,7698          12,6962          -2,424      0,0154      **
N_materiale     -1,47792          0,279666          -5,285      1,26e-07      ***
diametro        -0,00475963       0,00241207        -1,973      0,0485      **
anno_posa        0,0173570          0,00681953         2,545      0,0109      **
pMax            0,00624102       0,0112183         0,5563      0,5780
velocita        -0,000992300          0,0336507        -0,02949     0,9765
flusso          -0,00129556          0,000836005       -1,550      0,1212
TramRoute        1,23623          0,403819         3,061      0,0022      ***

Media var. dipendente  0,172619      SQM var. dipendente  0,378481
Log-verosimiglianza   -109,5921      Criterio di Akaike   235,1842
Criterio di Schwarz    265,7211      Hannan-Quinn        247,3570
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 290 (86,3%)
Test del rapporto di verosimiglianza: Chi-quadro(7) = 89,9458 [0,0000]
```

Figura 0-1 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM2-80%

L'indicatore visivo, così come i valori del *p-value*, segnalano la presenza di variabili poco significative. L'applicazione del metodo 'backward' riduce il modello, non rimuovendo soltanto le variabili statisticamente rilevanti.

```

Eliminazione sequenziale usando alfa (a due code) = 0,05

Scartata velocitA      (p-value 0,976)
Scartata pMax          (p-value 0,555)
Scartata flusso        (p-value 0,089)

Test sul modello 1

Ipotesi nulla: i parametri della regressione valgono zero per le variabili
pMax, velocitA, flusso
Statistica test: F(3, 328) = 1,05813, p-value 0,367035
L'omissione delle variabili ha migliorato 3 dei 3 criteri di informazione.

Modello 2: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

-----
              coefficiente      errore std.      z      p-value
-----
flag = 1
const          -29,1644          12,6099          -2,313    0,0207  **
N_materiale    -1,45405          0,276870          -5,252    1,51e-07  ***
diametro       -0,00534547         0,00252460         -2,117    0,0342  **
anno_posa       0,0164601          0,00673377         2,444    0,0145  **
TramRoute       1,01474           0,366030          2,772    0,0056  ***

Media var. dipendente  0,172619      SQM var. dipendente  0,378481
Log-verosimiglianza   -111,3357      Criterio di Akaike   232,6715
Criterio di Schwarz    251,7570      Hannan-Quinn        240,2795
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 296 (88,1%)
Test del rapporto di verosimiglianza: Chi-quadro(4) = 86,4585 [0,0000]

```

Figura 0-2 Stime dei parametri del modello applicato al set di dati Estrazione_TGM2-80% con variabili significative

Il modello di regressione logistica multivariata, in relazione ai risultati ottenuti, assume la forma:

$$\text{logit}[P(Y = 1)] = -29.1644 - 1.4540x_1 - 0.0053x_2 + 0.0164x_3 + 1.0147x_7$$

Ottenuti i coefficienti di regressione, la fase di test del modello viene eseguita sul restante 20% dei tratti. La probabilità di rottura per la generica condotta viene calcolata come:

$$\hat{\pi}(x) = \frac{e^{-29.1644 - 1.4540x_1 - 0.0053x_2 + 0.0164x_3 + 1.0147x_7}}{1 + e^{-29.1644 - 1.4540x_1 - 0.0053x_2 + 0.0164x_3 + 1.0147x_7}}$$

Estrazione TGM3-80%

Nell'output è possibile confrontare i risultati ottenuti applicando il modello di regressione, utilizzando il software *Gretl*:

Modello 1: Logit multinomiale, usando le osservazioni 1-336
 Variabile dipendente: flag
 Errori standard basati sull'Hessiana

	coefficiente	errore std.	z	p-value

flag = 1				
const	-7,68777	12,7363	-0,6036	0,5461
N_materiale	-1,25301	0,290459	-4,314	1,60e-05 ***
diametro	-0,00608112	0,00274609	-2,214	0,0268 **
anno_posa	0,00468158	0,00682406	0,6860	0,4927
pMax	0,0117184	0,0104428	1,122	0,2618
velocitA	0,0147270	0,0387069	0,3805	0,7036
flusso	0,000528875	0,000805026	0,6570	0,5112
TramRoute	1,19629	0,401388	2,980	0,0029 ***
Media var. dipendente	0,169643	SQM var. dipendente	0,375879	
Log-verosimiglianza	-102,2658	Criterio di Akaike	220,5317	
Criterio di Schwarz	251,0686	Hannan-Quinn	232,7046	

Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 299 (89,0%)
 Test del rapporto di verosimiglianza: Chi-quadro(7) = 101,443 [0,0000]

Figura 0-3 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM3-80%

L'indicatore visivo, così come i valori del *p-value*, segnalano la presenza di variabili poco significative. L'applicazione del metodo '*backward*' riduce il modello, considerando soltanto le variabili significative.

Eliminazione sequenziale usando alfa (a due code) = 0,05

Scartata velocitA	(p-value 0,704)
Scartata anno_posa	(p-value 0,467)
Scartata flusso	(p-value 0,270)
Scartata pMax	(p-value 0,182)

Test sul modello 1

Ipotesi nulla: i parametri della regressione valgono zero per le variabili
 anno_posa, pMax, velocitA, flusso
 Statistica test: F(4, 328) = 0,841596, p-value 0,499574
 L'omissione delle variabili ha migliorato 3 dei 3 criteri di informazione.

Modello 2: Logit multinomiale, usando le osservazioni 1-336
 Variabile dipendente: flag
 Errori standard basati sull'Hessiana

	coefficiente	errore std.	z	p-value

flag = 1				
const	1,91015	0,561815	3,400	0,0007 ***
N_materiale	-1,03360	0,174639	-5,918	3,25e-09 ***
diametro	-0,00555662	0,00261897	-2,122	0,0339 **
TramRoute	1,19311	0,359733	3,317	0,0009 ***
Media var. dipendente	0,169643	SQM var. dipendente	0,375879	
Log-verosimiglianza	-103,7954	Criterio di Akaike	215,5907	
Criterio di Schwarz	230,8592	Hannan-Quinn	221,6772	

Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 301 (89,6%)
 Test del rapporto di verosimiglianza: Chi-quadro(3) = 98,3839 [0,0000]

Figura 0-4 Stime dei parametri del modello applicato al set di dati Estrazione_TGM3-80% con variabili significative

Il modello di regressione logistica multivariata, in relazione ai risultati ottenuti, assume la forma:

$$\text{logit}[P(Y = 1)] = 1.9101 - 1.0336x_1 - 0.0055x_2 + 1.1931x_7$$

Ottenuti i coefficienti di regressione, la fase di test del modello viene eseguita sul restante 20% dei tratti.

La probabilità di rottura per la generica condotta viene calcolata come:

$$\hat{\pi}(x) = \frac{e^{1.9101-1.0336x_1-0.0055x_2+1.1931x_7}}{1 + e^{1.9101-1.0336x_1-0.0055x_2+1.1931x_7}}$$

Estrazione TGM4-80%

Si riporta l'output fornito dal software *Gretl*, che fornisce la stima dei parametri del modello logistico:

```
Modello 1: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana
```

	coefficiente	errore std.	z	p-value

flag = 1				
const	4,41057	11,2101	0,3934	0,6940
N_materiale	-1,05251	0,243601	-4,321	1,56e-05 ***
diametro	-0,00585505	0,00328970	-1,780	0,0751 *
anno_posa	-0,000952931	0,00594150	-0,1604	0,8726
pMax	0,00460154	0,00786360	0,5852	0,5584
velocita	-0,0232678	0,0374172	-0,6218	0,5340
flusso	-0,000145981	0,000869756	-0,1678	0,8667
TramRoute	0,780299	0,402358	1,939	0,0525 *
Media var. dipendente	0,172619	SQM var. dipendente	0,378481	
Log-verosimiglianza	-101,7049	Criterio di Akaike	219,4099	
Criterio di Schwarz	249,9468	Hannan-Quinn	231,5827	

Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 297 (88,4%)

Test del rapporto di verosimiglianza: Chi-quadro(7) = 105,72 [0,0000]

Figura 0-5 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM4-80%

L'indicatore visivo, così come i valori del *p-value*, segnalano la presenza di variabili poco significative. L'applicazione del metodo '*backward*' riduce il modello, considerando soltanto le variabili significative.


```

Eliminazione sequenziale usando alfa (a due code) = 0,05

Scartata anno_posa      (p-value 0,873)
Scartata flusso         (p-value 0,882)
Scartata pMax           (p-value 0,526)
Scartata velocitA       (p-value 0,498)
Scartata diametro       (p-value 0,054)

Test sul modello 1

Ipotesi nulla: i parametri della regressione valgono zero per le variabili
diametro, anno_posa, pMax, velocitA, flusso
Statistica test: F(5, 328) = 0,898012, p-value 0,482572
L'omissione delle variabili ha migliorato 3 dei 3 criteri di informazione.

Modello 2: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

-----
                coefficiente    errore std.      z      p-value
-----
flag = 1
const          1,95684         0,511825      3,823   0,0001   ***
N_materiale    -1,22810         0,169848     -7,231  4,81e-013 ***
TramRoute       0,742105        0,355122      2,090   0,0366   **

Media var. dipendente  0,172619    SQM var. dipendente  0,378481
Log-verosimiglianza   -104,9798    Criterio di Akaike   215,9595
Criterio di Schwarz   227,4109    Hannan-Quinn        220,5244
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 294 (87,5%)
Test del rapporto di verosimiglianza: Chi-quadro(2) = 99,1704 [0,0000]

```

Figura 0-6 Stime dei parametri del modello applicato al set di dati Estrazione_TGM4-80% con variabili significative

Il modello di regressione logistica multivariata, in relazione ai risultati ottenuti, assume la forma:

$$\text{logit}[P(Y = 1)] = 1.9568 - 1.2281x_1 + 0.7421x_7$$

Ottenuti i coefficienti di regressione, la fase di test del modello viene eseguita sul restante 20% dei tratti. La probabilità di rottura per la generica condotta viene calcolata come:

$$\hat{\pi}(x) = \frac{e^{1.9568 - 1.2281x_1 + 0.7421x_7}}{1 + e^{1.9568 - 1.2281x_1 + 0.7421x_7}}$$

Estrazione TGM5-80%

Si riporta l'output fornito dal software *Gretl* prendendo in considerazione tutte le variabili esplicative.


```

Modello 1: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

-----
               coefficiente   errore std.      z      p-value
-----
flag = 1
const          0,709942      13,3034      0,05337  0,9574
N_materiale    -1,15682      0,291425    -3,970   7,20e-05 ***
diametro       -0,00305196      0,00279039  -1,094   0,2741
anno_posa      0,000302085    0,00713799   0,04232  0,9662
pMax           0,00448182    0,00688096   0,6513   0,5148
velocitA       0,0225227      0,0391167    0,5758   0,5648
flusso         0,000197208    0,000822398  0,2398   0,8105
TramRoute      0,922533      0,420634     2,193    0,0283 **

Media var. dipendente  0,163690  SQM var. dipendente  0,370546
Log-verosimiglianza   -100,3147  Criterio di Akaike   216,6294
Criterio di Schwarz    247,1662  Hannan-Quinn         228,8022
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 296 (88,1%)
Test del rapporto di verosimiglianza: Chi-quadro(7) = 98,9074 [0,0000]

```

Figura 0-7 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM5-80%

L'indicatore visivo, così come i valori del *p-value*, segnalano la presenza di variabili poco significative. L'applicazione del metodo '*backward*' riduce il modello, non rimuovendo soltanto le variabili rilevanti.

```

Eliminazione sequenziale usando alfa (a due code) = 0,05

Scartata anno_posa      (p-value 0,966)
Scartata flusso         (p-value 0,806)
Scartata pMax           (p-value 0,545)
Scartata diametro       (p-value 0,238)
Scartata velocitA       (p-value 0,304)

Test sul modello 1

Ipotesi nulla: i parametri della regressione valgono zero per le variabili
diametro, anno_posa, pMax, velocitA, flusso
Statistica test: F(5, 328) = 0,674149, p-value 0,64332
L'omissione delle variabili ha migliorato 3 dei 3 criteri di informazione.

Modello 2: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

-----
               coefficiente   errore std.      z      p-value
-----
flag = 1
const          1,73918      0,528797     3,289    0,0010 ***
N_materiale    -1,18323      0,173241    -6,830    8,49e-012 ***
TramRoute      0,884884      0,361136     2,450    0,0143 **

Media var. dipendente  0,163690  SQM var. dipendente  0,370546
Log-verosimiglianza   -101,8968  Criterio di Akaike   209,7936
Criterio di Schwarz    221,2449  Hannan-Quinn         214,3584
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 295 (87,8%)
Test del rapporto di verosimiglianza: Chi-quadro(2) = 95,7431 [0,0000]

```

Figura 0-8 Stime dei parametri del modello applicato al set di dati Estrazione_TGM5-80% con variabili significative

Il modello di regressione logistica multivariata, in relazione ai risultati ottenuti, assume la forma:

$$\text{logit}[P(Y = 1)] = 1.7391 - 1.1832x_1 + 0.8848x_7$$

Ottenuti i coefficienti di regressione, la fase di test del modello viene eseguita sul restante 20% dei tratti. La probabilità di rottura per la generica condotta viene calcolata come:

$$\hat{\pi}(x) = \frac{e^{1.7391-1.1832x_1+0.8848x_7}}{1 + e^{1.7391-1.1832x_1+0.8848x_7}}$$

Estrazione TGM6-80%

Si riporta l'output fornito dal software *Gretl* prendendo in considerazione tutte le variabili esplicative.

```
Modello 2: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana
```

	coefficiente	errore std.	z	p-value
flag = 1				
const	-15,9484	12,5178	-1,274	0,2026
N_materiale	-1,43449	0,303711	-4,723	2,32e-06 ***
diametro	-0,00593824	0,00276966	-2,144	0,0320 **
anno_posa	0,0104021	0,00676451	1,538	0,1241
pMax	-0,0381832	0,0241272	-1,583	0,1135
velocita	0,0358732	0,0436578	0,8217	0,4113
flusso	-0,00119376	0,000959719	-1,244	0,2136
TramRoute	1,45798	0,442015	3,298	0,0010 ***
Media var. dipendente	0,157738	SQM var. dipendente	0,365039	
Log-verosimiglianza	-96,90060	Criterio di Akaike	209,8012	
Criterio di Schwarz	240,3381	Hannan-Quinn	221,9741	

Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 297 (88,4%)

Test del rapporto di verosimiglianza: Chi-quadro(7) = 99,1236 [0,0000]

Figura 0-9 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM6-80%

L'indicatore visivo, così come i valori del *p-value*, segnalano la presenza di variabili poco significative. L'applicazione del metodo 'backward' riduce il modello, non rimuovendo soltanto le variabili statisticamente rilevanti.

```

Eliminazione sequenziale usando alfa (a due code) = 0,05

Scartata velocita      (p-value 0,411)
Scartata flusso        (p-value 0,295)
Scartata pMax          (p-value 0,155)
Scartata anno_posa     (p-value 0,156)

Test sul modello 2

Ipotesi nulla: i parametri della regressione valgono zero per le variabili
anno_posa, pMax, velocita, flusso
Statistica test: F(4, 328) = 1,39117, p-value 0,236688
L'omissione delle variabili ha migliorato 3 dei 3 criteri di informazione.

Modello 5: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

-----
              coefficiente      errore std.      z      p-value
-----
flag = 1
const          1,91903          0,609324          3,149    0,0016 ***
N_materiale    -1,04686          0,193770         -5,403    6,57e-08 ***
diametro       -0,00644751        0,00284549        -2,266    0,0235 **
TramRoute       0,993324          0,363887          2,730    0,0063 ***

Media var. dipendente  0,157738      SQM var. dipendente  0,365039
Log-verosimiglianza   -100,0652      Criterio di Akaike   208,1304
Criterio di Schwarz    223,3989      Hannan-Quinn        214,2169
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 290 (86,3%)
Test del rapporto di verosimiglianza: Chi-quadro(3) = 92,7944 [0,0000]

```

Figura 0-10 Stime dei parametri del modello applicato al set di dati Estrazione_TGM6-80% con variabili significative

Il modello di regressione logistica multivariata, in relazione ai risultati ottenuti, assume la forma:

$$\text{logit}[P(Y = 1)] = 1.9190 - 1.0468x_1 - 0.0064x_2 + 0.9933x_7$$

Ottenuti i coefficienti di regressione, la fase di test del modello viene eseguita sul restante 20% dei tratti. La probabilità di rottura per la generica condotta viene calcolata come:

$$\hat{\pi}(x) = \frac{e^{1.9190 - 1.0468x_1 - 0.0064x_2 + 0.9933x_7}}{1 + e^{1.9190 - 1.0468x_1 - 0.0064x_2 + 0.9933x_7}}$$

Estrazione TGM7-80%

Nell'output è possibile confrontare i risultati ottenuti applicando il modello di regressione, utilizzando il software *Gretl*:

```

Modello 1: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

      coefficiente    errore std.      z      p-value
-----
flag = 1
const      -9,25520      14,8481      -0,6233    0,5331
N_materiale -0,556385     0,293230     -1,897     0,0578  *
diametro   -0,00793515     0,00439156   -1,807     0,0708  *
anno_posa   0,00468948     0,00785287    0,5972     0,5504
pMax        0,00265414     0,00857032    0,3097     0,7568
velocitaA   0,0120194      0,0378785     0,3173     0,7510
flusso      0,000232604    0,000945352   0,2461     0,8056
TramRoute   -0,177464      0,549906     -0,3227     0,7469

Media var. dipendente  0,068452    SQM var. dipendente  0,252897
Log-verosimiglianza    -74,60765    Criterio di Akaike    165,2153
Criterio di Schwarz    195,7522    Hannan-Quinn          177,3882
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 313 (93,2%)
Test del rapporto di verosimiglianza: Chi-quadro(7) = 18,5275 [0,0098]

```

Figura 0-11 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM7-80%

L'indicatore visivo, così come i valori del *p-value*, segnalano la presenza di variabili poco significative. L'applicazione del metodo '*backward*' riduce il modello, considerando soltanto le variabili significative.

```

Eliminazione sequenziale usando alfa (a due code) = 0,05

Scartata flusso          (p-value 0,806)
Scartata pMax            (p-value 0,788)
Scartata TramRoute       (p-value 0,791)
Scartata velocitaA       (p-value 0,523)
Scartata anno_posa       (p-value 0,452)
Scartata diametro        (p-value 0,070)

Test sul modello 1

Ipotesi nulla: i parametri della regressione valgono zero per le variabili
diametro, anno_posa, pMax, velocitaA, flusso, TramRoute
Statistica test: F(6, 328) = 0,732559, p-value 0,623692
L'omissione delle variabili ha migliorato 3 dei 3 criteri di informazione.

Modello 2: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

      coefficiente    errore std.      z      p-value
-----
flag = 1
const      -0,598482     0,609473     -0,9820     0,3261
N_materiale -0,570819     0,182556     -3,127     0,0018  ***

Media var. dipendente  0,068452    SQM var. dipendente  0,252897
Log-verosimiglianza    -78,02043    Criterio di Akaike    160,0409
Criterio di Schwarz    167,6751    Hannan-Quinn          163,0841
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 313 (93,2%)
Test del rapporto di verosimiglianza: Chi-quadro(1) = 11,7019 [0,0006]

```

Figura 0-12 Stime dei parametri del modello applicato al set di dati Estrazione_TGM7-80% con variabili significative

Il modello di regressione logistica multivariata, in relazione ai risultati ottenuti, assume la forma:

$$\text{logit}[P(Y = 1)] = -0.5708x_1$$

Ottenuti i coefficienti di regressione, la fase di test del modello viene eseguita sul restante 20% dei tratti.

La probabilità di rottura per la generica condotta viene calcolata come:

$$\hat{\pi}(x) = \frac{e^{-0.5708x_1}}{1 + e^{-0.5708x_1}}$$

Estrazione TGM8-80%

Si riporta l'output fornito dal software *Gretl*, che fornisce la stima dei parametri del modello logistico:

```
Modello 1: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana
```

	coefficiente	errore std.	z	p-value

flag = 1				
const	-14,7766	13,2624	-1,114	0,2652
N_materiale	-1,36321	0,300047	-4,543	5,54e-06 ***
diametro	-0,00483699	0,00300904	-1,607	0,1079
anno_posa	0,00853313	0,00711149	1,200	0,2302
pMax	0,00314694	0,00892801	0,3525	0,7245
velocita	0,0288430	0,0328890	0,8770	0,3805
flusso	-9,20836e-05	0,000881642	-0,1044	0,9168
TramRoute	1,38232	0,438329	3,154	0,0016 ***
Media var. dipendente	0,157738	SQM var. dipendente		0,365039
Log-verosimiglianza	-96,65149	Criterio di Akaike		209,3030
Criterio di Schwarz	239,8399	Hannan-Quinn		221,4758

Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 304 (90,5%)

Test del rapporto di verosimiglianza: Chi-quadro(7) = 99,6218 [0,0000]

Figura 0-13 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM8-80%

L'indicatore visivo, così come i valori del *p-value*, segnalano la presenza di variabili poco significative. L'applicazione del metodo '*backward*' riduce il modello, considerando soltanto le variabili significative.

```

Eliminazione sequenziale usando alfa (a due code) = 0,05

Scartata flusso          (p-value 0,917)
Scartata pMax            (p-value 0,711)
Scartata velocitA        (p-value 0,266)
Scartata anno_posa       (p-value 0,183)
Scartata diametro        (p-value 0,118)

Test sul modello 1

Ipotesi nulla: i parametri della regressione valgono zero per le variabili
diametro, anno_posa, pMax, velocitA, flusso
Statistica test: F(5, 328) = 1,16017, p-value 0,328591
L'omissione delle variabili ha migliorato 3 dei 3 criteri di informazione.

Modello 3: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

-----
              coefficiente      errore std.      z      p-value
-----
flag = 1
const          1,58593          0,534817      2,965    0,0030    ***
N_materiale    -1,17250          0,174306     -6,727    1,74e-011  ***
TramRoute       1,05233          0,365412      2,880    0,0040    ***

Media var. dipendente  0,157738      SQM var. dipendente  0,365039
Log-verosimiglianza   -99,75939      Criterio di Akaike   205,5188
Criterio di Schwarz   216,9701      Hannan-Quinn        210,0836
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 298 (88,7%)
Test del rapporto di verosimiglianza: Chi-quadro(2) = 93,406 [0,0000]

```

Figura 0-14 Stime dei parametri del modello applicato al set di dati Estrazione_TGM8-80% con variabili significative

Il modello di regressione logistica multivariata, in relazione ai risultati ottenuti, assume la forma:

$$\text{logit}[P(Y = 1)] = 1.5859 - 1.1725x_1 + 1.0523x_7$$

Ottenuti i coefficienti di regressione, la fase di test del modello viene eseguita sul restante 20% dei tratti. La probabilità di rottura per la generica condotta viene calcolata come:

$$\hat{\pi}(x) = \frac{e^{1.5859 - 1.1725x_1 + 1.0523x_7}}{1 + e^{1.5859 - 1.1725x_1 + 1.0523x_7}}$$

Estrazione TGM9-80%

Si riporta l'output fornito dal software *Gretl* prendendo in considerazione tutte le variabili esplicative.

```

Modello 1: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

      coefficiente    errore std.      z      p-value
-----
flag = 1
const      -9,45996      13,4794      -0,7018    0,4828
N_materiale -1,57526      0,315808     -4,988     6,10e-07 ***
diametro    -0,00494881     0,00274569    -1,802     0,0715 *
anno_posa    0,00609058     0,00725094     0,8400     0,4009
pMax         0,00331905     0,00972207     0,3414     0,7328
velocita     0,0285668      0,0489496     0,5836     0,5595
flusso       0,000601290    0,00106514     0,5645     0,5724
TramRoute    0,999561      0,476018      2,100      0,0357 **

Media var. dipendente 0,130952    SQM var. dipendente 0,337851
Log-verosimiglianza -83,51950    Criterio di Akaike 183,0390
Criterio di Schwarz 213,5759    Hannan-Quinn 195,2119
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 307 (91,4%)
Test del rapporto di verosimiglianza: Chi-quadro(7) = 93,8268 [0,0000]

```

Figura 0-15 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM9-80%

L'indicatore visivo, così come i valori del *p-value*, segnalano la presenza di variabili poco significative.

L'applicazione del metodo '*backward*' riduce il modello, non rimuovendo soltanto le variabili rilevanti.

```

Eliminazione sequenziale usando alfa (a due code) = 0,05

Scartata pMax      (p-value 0,733)
Scartata flusso    (p-value 0,624)
Scartata anno_posa (p-value 0,370)
Scartata velocita  (p-value 0,098)
Scartata diametro  (p-value 0,085)
Scartata TramRoute (p-value 0,071)

Test sul modello 1

Ipotesi nulla: i parametri della regressione valgono zero per le variabili
diametro, anno_posa, pMax, velocita, flusso, TramRoute
Statistica test: F(6, 328) = 1,70669, p-value 0,118688
L'omissione delle variabili ha migliorato 3 dei 3 criteri di informazione.

Modello 2: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

      coefficiente    errore std.      z      p-value
-----
flag = 1
const      2,60007      0,548942      4,737     2,17e-06 ***
N_materiale -1,38982      0,191838     -7,245     4,33e-013 ***

Media var. dipendente 0,130952    SQM var. dipendente 0,337851
Log-verosimiglianza -88,93049    Criterio di Akaike 181,8610
Criterio di Schwarz 189,4952    Hannan-Quinn 184,9042
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 296 (88,1%)
Test del rapporto di verosimiglianza: Chi-quadro(1) = 83,0048 [0,0000]

```

Figura 0-16 Stime dei parametri del modello applicato al set di dati Estrazione_TGM9-80% con variabili significative

Il modello di regressione logistica multivariata, in relazione ai risultati ottenuti, assume la forma:

$$\text{logit}[P(Y = 1)] = 2.6000 - 1.3898x_1$$

Ottenuti i coefficienti di regressione, la fase di test del modello viene eseguita sul restante 20% dei tratti.

La probabilità di rottura per la generica condotta viene calcolata come:

$$\hat{\pi}(x) = \frac{e^{2.6000 - 1.3898x_1}}{1 + e^{2.6000 - 1.3898x_1}}$$

Estrazione TGM10-80%

Si riporta l'output fornito dal software *Gretl* prendendo in considerazione tutte le variabili esplicative.

```
Modello 1: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana
```

	coefficiente	errore std.	z	p-value

flag = 1				
const	-7,47210	11,9604	-0,6247	0,5321
N_materiale	-1,21310	0,283908	-4,273	1,93e-05 ***
diametro	-0,00501085	0,00352524	-1,421	0,1552
anno_posa	0,00440238	0,00640263	0,6876	0,4917
pMax	-0,000871679	0,0102089	-0,08538	0,9320
velocita	0,0576078	0,0479806	1,201	0,2299
flusso	4,70841e-05	0,000937772	0,05021	0,9600
TramRoute	0,656430	0,458708	1,431	0,1524
Media var. dipendente	0,125000	SQM var. dipendente		0,331212
Log-verosimiglianza	-93,79538	Criterio di Akaike		203,5908
Criterio di Schwarz	234,1276	Hannan-Quinn		215,7636

Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 298 (88,7%)

Test del rapporto di verosimiglianza: Chi-quadro(7) = 65,5988 [0,0000]

Figura 0-17 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM10-80%

L'indicatore visivo, così come i valori del *p-value*, segnalano la presenza di variabili poco significative. L'applicazione del metodo 'backward' riduce il modello, non rimuovendo soltanto le variabili statisticamente rilevanti.


```

Eliminazione sequenziale usando alfa (a due code) = 0,05

Scartata flusso          (p-value 0,960)
Scartata pMax            (p-value 0,918)
Scartata anno_posa       (p-value 0,493)
Scartata TramRoute       (p-value 0,153)
Scartata diametro        (p-value 0,182)
Scartata velocitA        (p-value 0,150)

Test sul modello 1

Ipotesi nulla: i parametri della regressione valgono zero per le variabili
diametro, anno_posa, pMax, velocitA, flusso, TramRoute
Statistica test: F(6, 328) = 1,0543, p-value 0,390047
L'omissione delle variabili ha migliorato 3 dei 3 criteri di informazione.

Modello 2: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

-----
                coefficiente    errore std.      z      p-value
-----
flag = 1
const          1,59995         0,508236      3,148    0,0016 ***
N_materiale    -1,12776         0,182078     -6,194    5,87e-010 ***

Media var. dipendente  0,125000    SQM var. dipendente  0,331212
Log-verosimiglianza   -97,16627    Criterio di Akaike   198,3325
Criterio di Schwarz   205,9668    Hannan-Quinn        201,3757
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 297 (88,4%)
Test del rapporto di verosimiglianza: Chi-quadro(1) = 58,857 [0,0000]

```

Figura 0-18 Stime dei parametri del modello applicato al set di dati Estrazione_TGM10-80% con variabili significative

Il modello di regressione logistica multivariata, in relazione ai risultati ottenuti, assume la forma:

$$\text{logit}[P(Y = 1)] = 1.5999 - 1.1277x_1$$

Ottenuti i coefficienti di regressione, la fase di test del modello viene eseguita sul restante 20% dei tratti. La probabilità di rottura per la generica condotta viene calcolata come:

$$\hat{\pi}(x) = \frac{e^{1.5999 - 1.1277x_1}}{1 + e^{1.5999 - 1.1277x_1}}$$

Estrazione TGM11-80%

Nell'output è possibile confrontare i risultati ottenuti applicando il modello di regressione, utilizzando il software *Gretl*:

```

Modello 1: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

      coefficiente    errore std.      z      p-value
-----
flag = 1
const      -5,95179      11,7680      -0,5058    0,6130
N_materiale -1,18581      0,255202     -4,647     3,38e-06 ***
diametro    -0,00511967      0,00286387   -1,788     0,0738 *
anno_posa    0,00419021      0,00629023    0,6661     0,5053
pMax         0,00477248      0,00755007    0,6321     0,5273
velocitaA    0,00908973      0,0393817     0,2308     0,8175
flusso       -0,000707403      0,000873351   -0,8100     0,4179
TramRoute    0,937134      0,419822      2,232     0,0256 **

Media var. dipendente 0,172619    SQM var. dipendente 0,378481
Log-verosimiglianza -105,6423    Criterio di Akaike 227,2846
Criterio di Schwarz 257,8215    Hannan-Quinn 239,4575
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 296 (88,1%)
Test del rapporto di verosimiglianza: Chi-quadro(7) = 97,8453 [0,0000]

```

Figura 0-19 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM11-80%

L'indicatore visivo, così come i valori del *p-value*, segnalano la presenza di variabili poco significative. L'applicazione del metodo '*backward*' riduce il modello, considerando soltanto le variabili significative.

```

Eliminazione sequenziale usando alfa (a due code) = 0,05

Scartata velocitaA      (p-value 0,817)
Scartata anno_posa      (p-value 0,489)
Scartata flusso         (p-value 0,479)
Scartata pMax           (p-value 0,367)
Scartata diametro       (p-value 0,072)
Scartata TramRoute      (p-value 0,072)

Test sul modello 1

Ipotesi nulla: i parametri della regressione valgono zero per le variabili
diametro, anno_posa, pMax, velocitaA, flusso, TramRoute
Statistica test: F(6, 328) = 1,40877, p-value 0,210488
L'omissione delle variabili ha migliorato 3 dei 3 criteri di informazione.

Modello 2: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

      coefficiente    errore std.      z      p-value
-----
flag = 1
const      2,21720      0,460255      4,817     1,45e-06 ***
N_materiale -1,24097      0,167143     -7,425     1,13e-013 ***

Media var. dipendente 0,172619    SQM var. dipendente 0,378481
Log-verosimiglianza -110,6316    Criterio di Akaike 225,2632
Criterio di Schwarz 232,8974    Hannan-Quinn 228,3064
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 287 (85,4%)
Test del rapporto di verosimiglianza: Chi-quadro(1) = 87,8667 [0,0000]

```

Figura 0-20 Stime dei parametri del modello applicato al set di dati Estrazione_TGM11-80% con variabili significative

Il modello di regressione logistica multivariata, in relazione ai risultati ottenuti, assume la forma:

$$\text{logit}[P(Y = 1)] = 2.2172 - 1.2409x_1$$

Ottenuti i coefficienti di regressione, la fase di test del modello viene eseguita sul restante 20% dei tratti.

La probabilità di rottura per la generica condotta viene calcolata come:

$$\hat{\pi}(x) = \frac{e^{2.2172-1.2409x_1}}{1 + e^{2.2172-1.2409x_1}}$$

Estrazione TGM12-80%

Si riporta l'output fornito dal software *Gretl*, che fornisce la stima dei parametri del modello logistico:

```
Modello 1: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana
```

	coefficiente	errore std.	z	p-value
flag = 1				
const	-15,4577	12,6204	-1,225	0,2206
N materiale	-1,26350	0,289920	-4,358	1,31e-05 ***
diametro	-0,00491909	0,00257560	-1,910	0,0561 *
anno_posa	0,00874213	0,00677811	1,290	0,1971
pMax	0,00339621	0,00810707	0,4189	0,6753
velocita	0,0290319	0,0441892	0,6570	0,5112
flusso	-6,84017e-06	0,000861686	-0,007938	0,9937
TramRoute	1,15285	0,385897	2,987	0,0028 ***
Media var. dipendente	0,166667	SQM var. dipendente		0,373234
Log-verosimiglianza	-106,6964	Criterio di Akaike		229,3928
Criterio di Schwarz	259,9297	Hannan-Quinn		241,5657

Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 301 (89,6%)

Test del rapporto di verosimiglianza: Chi-quadro(7) = 89,3843 [0,0000]

Figura 0-21 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM12-80%

L'indicatore visivo, così come i valori del *p-value*, segnalano la presenza di variabili poco significative. L'applicazione del metodo '*backward*' riduce il modello, considerando soltanto le variabili significative.

```

Eliminazione sequenziale usando alfa (a due code) = 0,05

Scartata flusso          (p-value 0,994)
Scartata pMax            (p-value 0,657)
Scartata velocitA        (p-value 0,305)
Scartata anno_posa       (p-value 0,138)

Test sul modello 1

Ipotesi nulla: i parametri della regressione valgono zero per le variabili
anno_posa, pMax, velocitA, flusso
Statistica test: F(4, 328) = 0,879939, p-value 0,476108
L'omissione delle variabili ha migliorato 3 dei 3 criteri di informazione.

Modello 2: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

-----
              coefficiente      errore std.      z      p-value
-----
flag = 1
const          1,49250          0,537305          2,778    0,0055 ***
N_materiale    -0,916323         0,175314         -5,227    1,73e-07 ***
diametro       -0,00493556       0,00245699       -2,009    0,0446 **
TramRoute       1,01626          0,347740          2,922    0,0035 ***

Media var. dipendente  0,166667      SQM var. dipendente  0,373234
Log-verosimiglianza   -108,4071      Criterio di Akaike    224,8142
Criterio di Schwarz   240,0826      Hannan-Quinn         230,9006
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 290 (86,3%)
Test del rapporto di verosimiglianza: Chi-quadro(3) = 85,9629 [0,0000]

```

Figura 0-22 Stime dei parametri del modello applicato al set di dati Estrazione_TGM12-80% con variabili significative

Il modello di regressione logistica multivariata, in relazione ai risultati ottenuti, assume la forma:

$$\text{logit}[P(Y = 1)] = 1.4925 - 0.9163x_1 - 0.0049x_2 + 1.0162x_7$$

Ottenuti i coefficienti di regressione, la fase di test del modello viene eseguita sul restante 20% dei tratti. La probabilità di rottura per la generica condotta viene calcolata come:

$$\hat{\pi}(x) = \frac{e^{1.4925 - 0.9163x_1 - 0.0049x_2 + 1.0162x_7}}{1 + e^{1.4925 - 0.9163x_1 - 0.0049x_2 + 1.0162x_7}}$$

Estrazione TGM13-80%

Si riporta l'output fornito dal software *Gretl* prendendo in considerazione tutte le variabili esplicative.


```

Modello 1: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

      coefficiente    errore std.      z      p-value
-----
flag = 1
const      -3,89473      12,2435      -0,3181    0,7504
N_materiale -1,29794      0,284249     -4,566     4,97e-06 ***
diametro    -0,00496760      0,00233452    -2,128     0,0333 **
anno_posa    0,00263124      0,00656716     0,4007     0,6887
pMax        -0,00419021      0,00800009    -0,5238     0,6004
velocitA     0,0852519      0,0436401     1,954     0,0508 *
flusso      -0,000470178      0,000840961   -0,5591     0,5761
TramRoute    1,46966      0,431255      3,408     0,0007 ***

Media var. dipendente 0,172619  SQM var. dipendente 0,378481
Log-verosimiglianza -97,54093  Criterio di Akaike 211,0819
Criterio di Schwarz 241,6188  Hannan-Quinn 223,2547
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 304 (90,5%)
Test del rapporto di verosimiglianza: Chi-quadro(7) = 114,048 [0,0000]

```

Figura 0-23 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM13-80%

L'indicatore visivo, così come i valori del *p-value*, segnalano la presenza di variabili poco significative. L'applicazione del metodo '*backward*' riduce il modello, non rimuovendo soltanto le variabili rilevanti.

```

Eliminazione sequenziale usando alfa (a due code) = 0,05

Scartata anno_posa      (p-value 0,689)
Scartata flusso         (p-value 0,594)
Scartata pMax           (p-value 0,687)

Test sul modello 1

Ipotesi nulla: i parametri della regressione valgono zero per le variabili
anno_posa, pMax, flusso
Statistica test: F(3, 328) = 0,198165, p-value 0,897608
L'omissione delle variabili ha migliorato 3 dei 3 criteri di informazione.

Modello 2: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

      coefficiente    errore std.      z      p-value
-----
flag = 1
const      1,04651      0,746001      1,403     0,1607
N_materiale -1,21374      0,196267     -6,184     6,24e-010 ***
diametro    -0,00508827      0,00233287    -2,181     0,0292 **
velocitA     0,0665580      0,0315549     2,109     0,0349 **
TramRoute    1,34313      0,391740      3,429     0,0006 ***

Media var. dipendente 0,172619  SQM var. dipendente 0,378481
Log-verosimiglianza -97,85380  Criterio di Akaike 205,7076
Criterio di Schwarz 224,7931  Hannan-Quinn 213,3156
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 299 (89,0%)
Test del rapporto di verosimiglianza: Chi-quadro(4) = 113,422 [0,0000]

```

Figura 0-24 Stime dei parametri del modello applicato al set di dati Estrazione_TGM13-80% con variabili significative

Il modello di regressione logistica multivariata, in relazione ai risultati ottenuti, assume la forma:

$$\text{logit}[P(Y = 1)] = -1.2137x_1 - 0.0050x_2 + 0.0665x_5 + 1.3431x_7$$

Ottenuti i coefficienti di regressione, la fase di test del modello viene eseguita sul restante 20% dei tratti. La probabilità di rottura per la generica condotta viene calcolata come:

$$\hat{\pi}(x) = \frac{e^{-1.2137x_1 - 0.0050x_2 + 0.0665x_5 + 1.3431x_7}}{1 + e^{-1.2137x_1 - 0.0050x_2 + 0.0665x_5 + 1.3431x_7}}$$

Estrazione TGM14-80%

Si riporta l'output fornito dal software *Gretl* prendendo in considerazione tutte le variabili esplicative.

```
Modello 1: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana
```

	coefficiente	errore std.	z	p-value	
flag = 1					
const	-6,11591	11,9375	-0,5123	0,6084	
N_materiale	-1,12122	0,250184	-4,482	7,41e-06	***
diametro	-0,00597552	0,00266700	-2,241	0,0251	**
anno_posa	0,00326461	0,00633260	0,5155	0,6062	
pMax	0,0481526	0,0180690	2,665	0,0077	***
velocita	-0,00250352	0,0410799	-0,06094	0,9514	
flusso	-0,000765081	0,000953921	-0,8020	0,4225	
TramRoute	0,954059	0,419086	2,277	0,0228	**
Media var. dipendente	0,172619	SQM var. dipendente	0,378481		
Log-verosimiglianza	-102,9924	Criterio di Akaike	221,9849		
Criterio di Schwarz	252,5218	Hannan-Quinn	234,1577		

Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 290 (86,3%)
 Test del rapporto di verosimiglianza: Chi-quadro(7) = 103,145 [0,0000]

Figura 0-25 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM14-80%

L'indicatore visivo, così come i valori del *p-value*, segnalano la presenza di variabili poco significative. L'applicazione del metodo 'backward' riduce il modello, non rimuovendo soltanto le variabili statisticamente rilevanti.

```

Eliminazione sequenziale usando alfa (a due code) = 0,05

Scartata velocitA      (p-value 0,951)
Scartata anno_posa     (p-value 0,609)
Scartata flusso        (p-value 0,366)

Test sul modello 1

Ipotesi nulla: i parametri della regressione valgono zero per le variabili
anno_posa, velocitA, flusso
Statistica test: F(3, 328) = 0,356037, p-value 0,784804
L'omissione delle variabili ha migliorato 3 dei 3 criteri di informazione.

Modello 2: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

-----
              coefficiente      errore std.      z      p-value
-----
flag = 1
const          -0,141062         0,909572      -0,1551    0,8768
N_materiale    -1,02541          0,175670     -5,837    5,31e-09 ***
diametro       -0,00631145         0,00267663    -2,358    0,0184 **
pMax           0,0457966         0,0169515     2,702    0,0069 ***
TramRoute      0,819351          0,368436     2,224    0,0262 **

Media var. dipendente 0,172619      SQM var. dipendente 0,378481
Log-verosimiglianza  -103,5520      Criterio di Akaike 217,1041
Criterio di Schwarz 236,1896      Hannan-Quinn 224,7121
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 290 (86,3%)
Test del rapporto di verosimiglianza: Chi-quadro(4) = 102,026 [0,0000]

```

Figura 0-26 Stime dei parametri del modello applicato al set di dati Estrazione_TGM14-80% con variabili significative

Il modello di regressione logistica multivariata, in relazione ai risultati ottenuti, assume la forma:

$$\text{logit}[P(Y = 1)] = -1.0254x_1 - 0.0063x_2 + 0.04579x_4 + 0.8193x_7$$

Ottenuti i coefficienti di regressione, la fase di test del modello viene eseguita sul restante 20% dei tratti. La probabilità di rottura per la generica condotta viene calcolata come:

$$\hat{\pi}(x) = \frac{e^{-1.0254x_1 - 0.0063x_2 + 0.04579x_4 + 0.8193x_7}}{1 + e^{-1.0254x_1 - 0.0063x_2 + 0.04579x_4 + 0.8193x_7}}$$

Estrazione TGM15-80%

Nell'output è possibile confrontare i risultati ottenuti applicando il modello di regressione, utilizzando il software *Gretl*:

```

Modello 1: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

      coefficiente      errore std.      z      p-value
-----
flag = 1
const      -14,0877      12,0679      -1,167      0,2431
N_materiale -1,29668      0,266996      -4,857      1,19e-06 ***
diametro    -0,00720833      0,00266199      -2,708      0,0068 ***
anno_posa    0,00785115      0,00649288      1,209      0,2266
pMax         0,0234436      0,0134761      1,740      0,0819 *
velocitA     0,0156025      0,0441796      0,3532      0,7240
flusso       0,000429727      0,000894418      0,4805      0,6309
TramRoute    1,56240      0,438933      3,560      0,0004 ***

Media var. dipendente 0,175595      SQM var. dipendente 0,381043
Log-verosimiglianza -96,89056      Criterio di Akaike 209,7811
Criterio di Schwarz 240,3180      Hannan-Quinn 221,9540
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 304 (90,5%)
Test del rapporto di verosimiglianza: Chi-quadro(7) = 118,462 [0,0000]

```

Figura 0-27 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM15-80%

L'indicatore visivo, così come i valori del *p-value*, segnalano la presenza di variabili poco significative.

L'applicazione del metodo '*backward*' riduce il modello, considerando soltanto le variabili significative.

```

Eliminazione sequenziale usando alfa (a due code) = 0,05

Scartata velocitA      (p-value 0,724)
Scartata flusso        (p-value 0,365)
Scartata anno_posa     (p-value 0,158)
Scartata pMax          (p-value 0,062)

Test sul modello 1

Ipotesi nulla: i parametri della regressione valgono zero per le variabili
anno_posa, pMax, velocitA, flusso
Statistica test: F(4, 328) = 1,56651, p-value 0,182907
L'omissione delle variabili ha migliorato 3 dei 3 criteri di informazione.

Modello 2: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

      coefficiente      errore std.      z      p-value
-----
flag = 1
const      1,99628      0,539198      3,702      0,0002 ***
N_materiale -0,977452      0,172265      -5,674      1,39e-08 ***
diametro    -0,00660089      0,00262908      -2,511      0,0120 **
TramRoute    1,50620      0,386804      3,894      9,86e-05 ***

Media var. dipendente 0,175595      SQM var. dipendente 0,381043
Log-verosimiglianza -99,97122      Criterio di Akaike 207,9424
Criterio di Schwarz 223,2109      Hannan-Quinn 214,0289
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 301 (89,6%)
Test del rapporto di verosimiglianza: Chi-quadro(3) = 112,301 [0,0000]

```

Figura 0-28 Stime dei parametri del modello applicato al set di dati Estrazione_TGM15-80% con variabili significative

Il modello di regressione logistica multivariata, in relazione ai risultati ottenuti, assume la forma:

$$\text{logit}[P(Y = 1)] = 1.9962 - 0.9774x_1 - 0.0066x_2 + 1.5062x_7$$

Ottenuti i coefficienti di regressione, la fase di test del modello viene eseguita sul restante 20% dei tratti.

La probabilità di rottura per la generica condotta viene calcolata come:

$$\hat{\pi}(x) = \frac{e^{1.9962-0.9774x_1-0.0066x_2+1.5062x_7}}{1 + e^{1.9962-0.9774x_1-0.0066x_2+1.5062x_7}}$$

Estrazione TGM16-80%

Si riporta l'output fornito dal software *Gretl*, che fornisce la stima dei parametri del modello logistico:

Modello 1: Logit multinomiale, usando le osservazioni 1-336				
Variabile dipendente: flag				
Errori standard basati sull'Hessiana				
	coefficiente	errore std.	z	p-value

flag = 1				
const	16,1546	15,2695	1,058	0,2901
N_materiale	-0,730565	0,314170	-2,325	0,0201 **
diametro	-0,00493304	0,00252854	-1,951	0,0511 *
anno_posa	-0,00802006	0,00815922	-0,9829	0,3256
pMax	0,00290882	0,00769734	0,3779	0,7055
velocita	0,0174344	0,0377248	0,4621	0,6440
flusso	0,000609029	0,000844264	0,7214	0,4707
TramRoute	0,626729	0,423903	1,478	0,1393
Media var. dipendente	0,145833	SQM var. dipendente	0,353465	
Log-verosimiglianza	-100,4359	Criterio di Akaike	216,8718	
Criterio di Schwarz	247,4087	Hannan-Quinn	229,0447	
Note: SQM = scarto quadratico medio; E.S. = errore standard				
Numero dei casi 'previsti correttamente' = 297 (88,4%)				
Test del rapporto di verosimiglianza: Chi-quadro(7) = 78,2857 [0,0000]				

Figura 0-29 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM16-80%

L'indicatore visivo, così come i valori del *p-value*, segnalano la presenza di variabili poco significative. L'applicazione del metodo '*backward*' riduce il modello, considerando soltanto le variabili significative.

```

Eliminazione sequenziale usando alfa (a due code) = 0,05

Scartata pMax          (p-value 0,706)
Scartata velocita      (p-value 0,543)
Scartata anno_posa     (p-value 0,319)
Scartata flusso        (p-value 0,292)
Scartata diametro      (p-value 0,077)

Test sul modello 1

Ipotesi nulla: i parametri della regressione valgono zero per le variabili
diametro, anno_posa, pMax, velocita, flusso
Statistica test: F(5, 328) = 1,12973, p-value 0,344317
L'omissione delle variabili ha migliorato 3 dei 3 criteri di informazione.

Modello 2: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

-----
              coefficiente      errore std.      z      p-value
-----
flag = 1
const          1,45800          0,551300          2,645    0,0082    ***
N_materiale    -1,08256          0,169649          -6,381    1,76e-010 ***
TramRoute       0,713146          0,360860          1,976    0,0481    **

Media var. dipendente  0,145833      SQM var. dipendente  0,353465
Log-verosimiglianza   -103,8718      Criterio di Akaike    213,7436
Criterio di Schwarz   225,1949      Hannan-Quinn         218,3084
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 295 (87,8%)
Test del rapporto di verosimiglianza: Chi-quadro(2) = 71,4139 [0,0000]

```

Figura 0-30 Stime dei parametri del modello applicato al set di dati Estrazione_TGM16-80% con variabili significative

Il modello di regressione logistica multivariata, in relazione ai risultati ottenuti, assume la forma:

$$\text{logit}[P(Y = 1)] = 1.4580 - 1.0825x_1 + 0.7131x_7$$

Ottenuti i coefficienti di regressione, la fase di test del modello viene eseguita sul restante 20% dei tratti. La probabilità di rottura per la generica condotta viene calcolata come:

$$\hat{\pi}(x) = \frac{e^{1.4580 - 1.0825x_1 + 0.7131x_7}}{1 + e^{1.4580 - 1.0825x_1 + 0.7131x_7}}$$

Estrazione TGM17-80%

Si riporta l'output fornito dal software *Gretl* prendendo in considerazione tutte le variabili esplicative.

```

Modello 1: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

      coefficiente      errore std.      z      p-value
-----
flag = 1
const      -14,6740      12,4606      -1,178      0,2389
N_materiale -1,30343      0,303492      -4,295      1,75e-05 ***
diametro    -0,00487583      0,00220838      -2,208      0,0273 **
anno_posa     0,00850843      0,00676641      1,257      0,2086
pMax         0,00410537      0,0288926      0,1421      0,8870
velocita     0,0232065      0,0413139      0,5617      0,5743
flusso      -0,000380212      0,000903478      -0,4208      0,6739
TramRoute     1,12935      0,442008      2,555      0,0106 **

Media var. dipendente 0,154762      SQM var. dipendente 0,362217
Log-verosimiglianza -100,5809      Criterio di Akaike 217,1619
Criterio di Schwarz 247,6988      Hannan-Quinn 229,3347
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 299 (89,0%)
Test del rapporto di verosimiglianza: Chi-quadro(7) = 88,3901 [0,0000]

```

Figura 0-31 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM17-80%

L'indicatore visivo, così come i valori del *p-value*, segnalano la presenza di variabili poco significative.

L'applicazione del metodo '*backward*' riduce il modello, non rimuovendo soltanto le variabili rilevanti.

```

Eliminazione sequenziale usando alfa (a due code) = 0,05

Scartata pMax      (p-value 0,887)
Scartata flusso    (p-value 0,642)
Scartata velocita  (p-value 0,704)
Scartata anno_posa (p-value 0,185)

Test sul modello 1

Ipotesi nulla: i parametri della regressione valgono zero per le variabili
anno_posa, pMax, velocita, flusso
Statistica test: F(4, 328) = 0,530802, p-value 0,713184
L'omissione delle variabili ha migliorato 3 dei 3 criteri di informazione.

Modello 2: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

      coefficiente      errore std.      z      p-value
-----
flag = 1
const      1,67097      0,558836      2,990      0,0028 ***
N_materiale -0,995517      0,185419      -5,369      7,92e-08 ***
diametro    -0,00481575      0,00217104      -2,218      0,0265 **
TramRoute     0,946131      0,364658      2,595      0,0095 ***

Media var. dipendente 0,154762      SQM var. dipendente 0,362217
Log-verosimiglianza -101,6715      Criterio di Akaike 211,3430
Criterio di Schwarz 226,6115      Hannan-Quinn 217,4295
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 294 (87,5%)
Test del rapporto di verosimiglianza: Chi-quadro(3) = 86,209 [0,0000]

```

Figura 0-32 Stime dei parametri del modello applicato al set di dati Estrazione_TGM17-80% con variabili significative

Il modello di regressione logistica multivariata, in relazione ai risultati ottenuti, assume la forma:

$$\text{logit}[P(Y = 1)] = 1.6709 - 0.9955x_1 - 0.0048x_2 + 0.9461x_7$$

Ottenuti i coefficienti di regressione, la fase di test del modello viene eseguita sul restante 20% dei tratti.

La probabilità di rottura per la generica condotta viene calcolata come:

$$\hat{\pi}(x) = \frac{e^{1.6709 - 0.9955x_1 - 0.0048x_2 + 0.9461x_7}}{1 + e^{1.6709 - 0.9955x_1 - 0.0048x_2 + 0.9461x_7}}$$

Estrazione TGM18-80%

Nell'output è possibile confrontare i risultati ottenuti applicando il modello di regressione, utilizzando il software *Gretl*:

```
Modello 1: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana
```

	coefficiente	errore std.	z	p-value
flag = 1				
const	-8,76502	13,3202	-0,6580	0,5105
N_materiale	-1,33531	0,301916	-4,423	9,74e-06 ***
diametro	-0,00492831	0,00247605	-1,990	0,0465 **
anno_posa	0,00563152	0,00714108	0,7886	0,4303
pMax	0,00211003	0,00770134	0,2740	0,7841
velocita	0,0161674	0,0365853	0,4419	0,6586
flusso	0,000299577	0,000854826	0,3505	0,7260
TramRoute	0,891566	0,401563	2,220	0,0264 **
Media var. dipendente	0,169643	SQM var. dipendente	0,375879	
Log-verosimiglianza	-102,4913	Criterio di Akaike	220,9826	
Criterio di Schwarz	251,5195	Hannan-Quinn	233,1555	

Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 296 (88,1%)

Test del rapporto di verosimiglianza: Chi-quadro(7) = 100,992 [0,0000]

Figura 0-33 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM18-80%

L'indicatore visivo, così come i valori del *p-value*, segnalano la presenza di variabili poco significative.

L'applicazione del metodo '*backward*' riduce il modello, considerando soltanto le variabili significative.


```

Eliminazione sequenziale usando alfa (a due code) = 0,05

Scartata pMax          (p-value 0,784)
Scartata flusso        (p-value 0,779)
Scartata anno_posa     (p-value 0,419)
Scartata velocitA      (p-value 0,219)
Scartata diametro      (p-value 0,058)

Test sul modello 1

Ipotesi nulla: i parametri della regressione valgono zero per le variabili
diametro, anno_posa, pMax, velocitA, flusso
Statistica test: F(5, 328) = 1,10621, p-value 0,356858
L'omissione delle variabili ha migliorato 3 dei 3 criteri di informazione.

Modello 2: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

-----
              coefficiente   errore std.      z      p-value
-----
flag = 1
const          1,85871       0,528355      3,518   0,0004   ***
N_materiale    -1,21922       0,173395     -7,031  2,04e-012 ***
TramRoute       0,741155      0,351748      2,107   0,0351   **

Media var. dipendente  0,169643   SQM var. dipendente  0,375879
Log-verosimiglianza   -105,9798   Criterio di Akaike   217,9596
Criterio di Schwarz   229,4109   Hannan-Quinn        222,5244
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 292 (86,9%)
Test del rapporto di verosimiglianza: Chi-quadro(2) = 94,0151 [0,0000]

```

Figura 0-34 Stime dei parametri del modello applicato al set di dati Estrazione_TGM18-80% con variabili significative

Il modello di regressione logistica multivariata, in relazione ai risultati ottenuti, assume la forma:

$$\text{logit}[P(Y = 1)] = 1.8587 - 1.2192x_1 + 0.7411x_7$$

Ottenuti i coefficienti di regressione, la fase di test del modello viene eseguita sul restante 20% dei tratti.

La probabilità di rottura per la generica condotta viene calcolata come:

$$\hat{\pi}(x) = \frac{e^{1.8587 - 1.2192x_1 + 0.7411x_7}}{1 + e^{1.8587 - 1.2192x_1 + 0.7411x_7}}$$

Estrazione TGM19-80%

Si riporta l'output fornito dal software *Gretl*, che fornisce la stima dei parametri del modello logistico:

Modello 1: Logit multinomiale, usando le osservazioni 1-336
 Variabile dipendente: flag
 Errori standard basati sull'Hessiana

	coefficiente	errore std.	z	p-value	

flag = 1					
const	-5,99219	12,1530	-0,4931	0,6220	
N_materiale	-1,32594	0,311246	-4,260	2,04e-05	***
diametro	-0,00482676	0,00267835	-1,802	0,0715	*
anno_posa	0,00491117	0,00665536	0,7379	0,4606	
pMax	-0,0295873	0,0285286	-1,037	0,2997	
velocitA	0,0221906	0,0487779	0,4549	0,6492	
flusso	-0,000419131	0,000990442	-0,4232	0,6722	
TramRoute	1,32280	0,431881	3,063	0,0022	***
Media var. dipendente	0,154762	SQM var. dipendente	0,362217		
Log-verosimiglianza	-93,72734	Criterio di Akaike	203,4547		
Criterio di Schwarz	233,9916	Hannan-Quinn	215,6275		
Note: SQM = scarto quadratico medio; E.S. = errore standard					
Numero dei casi 'previsti correttamente' = 297 (88,4%)					
Test del rapporto di verosimiglianza: Chi-quadro(7) = 102,097 [0,0000]					

Figura 0-35 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM19-80%

L'indicatore visivo, così come i valori del *p-value*, segnalano la presenza di variabili poco significative.

L'applicazione del metodo 'backward' riduce il modello, considerando soltanto le variabili significative.

Eliminazione sequenziale usando alfa (a due code) = 0,05

Scartata flusso	(p-value 0,672)
Scartata velocitA	(p-value 0,784)
Scartata anno_posa	(p-value 0,453)
Scartata pMax	(p-value 0,413)

Test sul modello 1

Ipotesi nulla: i parametri della regressione valgono zero per le variabili anno_posa, pMax, velocitA, flusso
 Statistica test: F(4, 328) = 0,360593, p-value 0,836585
 L'omissione delle variabili ha migliorato 3 dei 3 criteri di informazione.

Modello 2: Logit multinomiale, usando le osservazioni 1-336
 Variabile dipendente: flag
 Errori standard basati sull'Hessiana

	coefficiente	errore std.	z	p-value	

flag = 1					
const	2,00799	0,582859	3,445	0,0006	***
N_materiale	-1,13076	0,196297	-5,760	8,39e-09	***
diametro	-0,00523191	0,00265925	-1,967	0,0491	**
TramRoute	1,14734	0,372820	3,077	0,0021	***
Media var. dipendente	0,154762	SQM var. dipendente	0,362217		
Log-verosimiglianza	-94,52057	Criterio di Akaike	197,0411		
Criterio di Schwarz	212,3096	Hannan-Quinn	203,1276		
Note: SQM = scarto quadratico medio; E.S. = errore standard					
Numero dei casi 'previsti correttamente' = 296 (88,1%)					
Test del rapporto di verosimiglianza: Chi-quadro(3) = 100,511 [0,0000]					

Figura 0-36 Stime dei parametri del modello applicato al set di dati Estrazione_TGM19-80% con variabili significative

Il modello di regressione logistica multivariata, in relazione ai risultati ottenuti, assume la forma:

$$\text{logit}[P(Y = 1)] = 2.0079 - 1.1307x_1 - 0.0052x_2 + 1.1473x_7$$

Ottenuti i coefficienti di regressione, la fase di test del modello viene eseguita sul restante 20% dei tratti.

La probabilità di rottura per la generica condotta viene calcolata come:

$$\hat{\pi}(x) = \frac{e^{2.0079-1.1307x_1-0.0052x_2+1.1473x_7}}{1 + e^{2.0079-1.1307x_1-0.0052x_2+1.1473x_7}}$$

Estrazione TGM20-80%

Si riporta l'output fornito dal software *Gretl* prendendo in considerazione tutte le variabili esplicative.

```
Modello 1: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana
```

	coefficiente	errore std.	z	p-value

flag = 1				
const	-4,20537	12,0052	-0,3503	0,7261
N_materiale	-1,19133	0,272159	-4,377	1,20e-05 ***
diametro	-0,00270475	0,00262506	-1,030	0,3028
anno_posa	0,00185429	0,00638200	0,2906	0,7714
pMax	0,0478063	0,0190627	2,508	0,0121 **
velocita	-0,00154181	0,0370728	-0,04159	0,9668
flusso	0,000344653	0,000947132	0,3639	0,7159
TramRoute	0,576506	0,440978	1,307	0,1911
Media var. dipendente	0,130952	SQM var. dipendente	0,337851	
Log-verosimiglianza	-95,54977	Criterio di Akaike	207,0995	
Criterio di Schwarz	237,6364	Hannan-Quinn	219,2724	

Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 294 (87,5%)
 Test del rapporto di verosimiglianza: Chi-quadro(7) = 69,7663 [0,0000]

Figura 0-37 Stime dei parametri del modello di regressione logistica applicato al set di dati Estrazione_TGM20-80%

L'indicatore visivo, così come i valori del *p-value*, segnalano la presenza di variabili poco significative. L'applicazione del metodo '*backward*' riduce il modello, non rimuovendo soltanto le variabili rilevanti.

```

Eliminazione sequenziale usando alfa (a due code) = 0,05

Scartata velocitA      (p-value 0,967)
Scartata anno_posa     (p-value 0,773)
Scartata flusso        (p-value 0,620)
Scartata diametro      (p-value 0,344)
Scartata TramRoute     (p-value 0,107)

Test sul modello 1

Ipotesi nulla: i parametri della regressione valgono zero per le variabili
diametro, anno_posa, velocitA, flusso, TramRoute
Statistica test: F(5, 328) = 0,745114, p-value 0,590173
L'omissione delle variabili ha migliorato 3 dei 3 criteri di informazione.

Modello 2: Logit multinomiale, usando le osservazioni 1-336
Variabile dipendente: flag
Errori standard basati sull'Hessiana

-----
                coefficiente    errore std.      z          p-value
-----
flag = 1
const          -0,623818        0,936795     -0,6659     0,5055
N_materiale    -1,20248         0,187273     -6,421      1,35e-010 ***
pMax           0,0483172        0,0182623     2,646      0,0082 ***

Media var. dipendente  0,130952    SQM var. dipendente  0,337851
Log-verosimiglianza   -97,50184    Criterio di Akaike   201,0037
Criterio di Schwarz   212,4550    Hannan-Quinn         205,5685
Note: SQM = scarto quadratico medio; E.S. = errore standard

Numero dei casi 'previsti correttamente' = 290 (86,3%)
Test del rapporto di verosimiglianza: Chi-quadro(2) = 65,8621 [0,0000]

```

Figura 0-38 Stime dei parametri del modello applicato al set di dati Estrazione_TGM20-80% con variabili significative

Il modello di regressione logistica multivariata, in relazione ai risultati ottenuti, assume la forma:

$$\text{logit}[P(Y = 1)] = -1.20248x_1 + 0.0483x_4$$

Ottenuti i coefficienti di regressione, la fase di test del modello viene eseguita sul restante 20% dei tratti.

La probabilità di rottura per la generica condotta viene calcolata come:

$$\hat{\pi}(x) = \frac{e^{-1.20248x_1 + 0.0483x_4}}{1 + e^{-1.20248x_1 + 0.0483x_4}}$$

BIBLIOGRAFIA

- Aslani, P. (2003). *Hazard Rate Modelling and Risk Analysis of Water Mains*. New York, USA: Polytechnic University.
- D.C., M., E., P., & G.G., V. (2012). *Introduction to Linear Regression Analysis*. John Wiley & Sons.
- Dell'Orfano, F. (2013). *Modello fisico-matematico per la rappresentazione del deterioramento delle condotte in un acquedotto*. Tesi di Dottorato. Napoli: Università degli Studi di Napoli Federico II.
- Ghigo, C. (2017). *Analisi dei guasti nella rete acquedottistica di Torino*. Tesi di Laurea Magistrale. Torino: Politecnico di Torino.
- Harrell FE, Lee KL, Mark DB. (1996) *Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors*. Statistics in Medicine.
- Kleiner, Y. Rajani, B. (2001). *Comprehensive review of structural deterioration of water mains: statistical models*. Urban water 3.
- Kleiner, Y., Rajani, B. (2002). *Modelling the deterioration of water mains and planning their renewal*. Montreal, Quebec: International Conference on Urban Infrastructure.
- Klersy, C., Scudeller, L. (2002). *Interpretare I modelli prognostici multivariati: il modello logistico*. G Ital Aritmol Cardestim, 4:205-209.
- Longo, A. (2011). *Analisi delle rotture in condotta: proposta di un criterio decisionale*. Tesi di Laurea Magistrale. Milano: Politecnico di Milano.
- Mantica, I. (1979). *Modello matematico per le reti di distribuzione degli acquedotti urbani ad anello*. il Giornale del Genio Civile.
- Normativa UNI. Consultato l'ultima volta a luglio 2019 da www.uni.com
- Parodi, S. (2007). *Applicazione del modello di regressione di Poisson e del modello logistico in Epidemiologia*.

- Saporito, G. (2018). *Analisi dei guasti nella rete acquedottistica di Torino*. Tesi di Laurea Magistrale. Torino: Politecnico di Torino.
- SMAT. Consultato l'ultima volta a settembre 2019 da www.smat.it
- Stephens, K., Tech, E., & Jackson, J. (s.d.). (2003). *Main Rehabilitation Prioritization*.
- Taufer, E. (2015). *Regressione logistica. Strumenti quantitativi per la gestione*.
- Tavernelli, M. (2011). *Analisi delle rotture nelle tubazioni delle reti acquedottistiche*. Tesi di Laurea Magistrale. Bologna: Università degli Studi di Bologna.
- Wengström, T. (1993). *Comparative Analysis of Pipe Break Rates. A Literature review*. Göteborg: Chalmers University of Technology.