Modelli per lo studio dell’incidentalità stradale

Lorenzo Mussone

Monza, 4 Aprile 2019

Indice della presentazione

- Introduzione
- I dati
- Gli indicatori
- I modelli analitici
- Osservazioni conclusive

Settori coinvolti nell’analisi degli incidenti stradali

- Conoscenza delle differenti infrastrutture stradali
- Conoscenza delle regole di circolazione
- Conoscenza del mezzo meccanico
- Valutazione delle possibili interazioni strada-veicolo
- Valutazione delle capacità o limiti del guidatore

Visione integrata del fenomeno traffico

- Geometria della strada e delle intersezioni
- Monitoraggio dei flussi
- Sorveglianza
- Regolazione dei flussi
- Pianificazione dei trasporti (e urbanistica)
- Sistemi di Informazione

Prestazioni veicolari

Misura delle prestazioni

↓

Modellazione delle interazioni tra utente-veicolo-strada-ambiente

↓

Derivazione di leggi da utilizzare a supporto di scelte progettuali del veicolo e della strada

- **Sicurezza:**
  - equilibrio
  - visibilità
  - manovra
- **Comfort:**
  - fisico
  - psicologico
- **Compatibilità e inserimento urbanistico**

Il traffico è un fenomeno aleatorio

- Non è esattamente prevedibile a livello microscopico per :
  - Flussi
  - Comportamenti
  - Scelte
- È legato a molti parametri :
  - Caratteristiche socio-economiche
  - Caratteristiche urbanistiche
  - Condizioni meteorologiche
- È una sommatoria di movimenti con motivazioni non necessariamente correlate

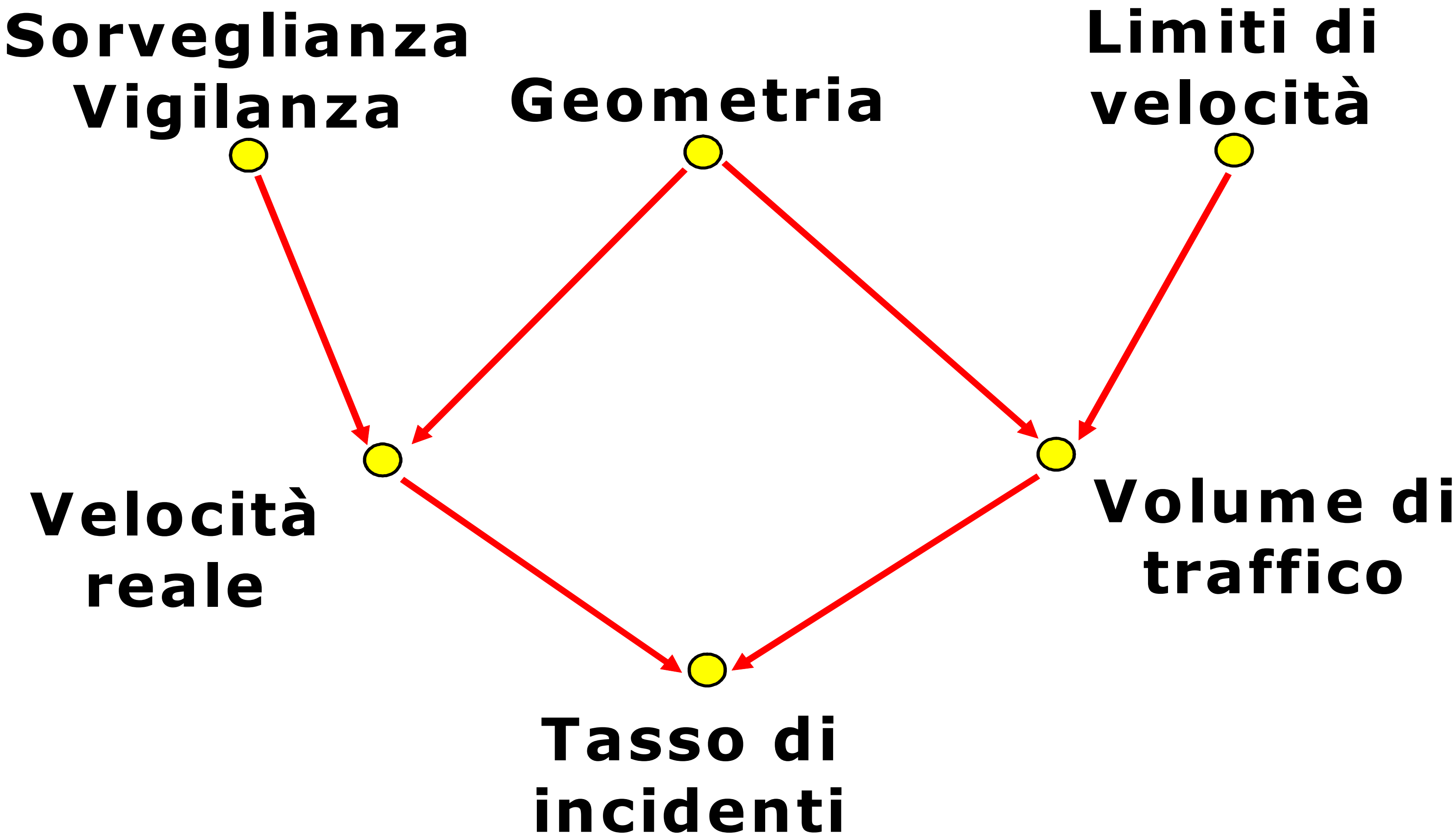
Le caratteristiche necessarie del rilevamento dati

- La significatività statistica dei dati deve essere sempre salvaguardata
- Gli incidenti sono un fenomeno raro: tutti gli incidenti hanno un contenuto informativo rilevante
- La pericolosità si manifesta indubbiamente tramite gli incidenti ma potrebbe richiedere altri stimatori/indicatori per una descrizione appropriata
- L’incidente non è un indicatore esaustivo della pericolosità

Sicurezza e incidenti veicolari

- La valutazione della sicurezza di un'infrastruttura stradale è una operazione complessa che deve tenere conto degli effetti non lineari dei parametri che entrano in gioco
- Cosa si intende per incidente? Accident o Crash?
- Disparità circa le modalità di rilevazione
- Frequenza (o tasso) di incidentalità
- Analisi del rischio

Diagramma delle variabili e relative relazioni del tasso di incidente



L'iceberg degli incidenti e del comportamento di guida

Linea di separazione tra incidente e non-incidente

Mortale

Grave

Lieve

Solo danni materiali

Appena mancato (conflitto)

Atto deliberatamente illegale

Errori di guida: infrazioni, distrazioni

Comportamenti di guida corretti: velocità, headway/gap, sorpasso,

Indicatori dell'incidentalità

- Sono di natura aggregata nello spazio e/o nel tempo
- Indicatori di gravità/frequenza
  - Per strada, per tipologia di strade, per rete
  - Per periodo di tempo (di solito l'anno)
  - Per tipo di veicolo, per utenza
  - Per percorrenza o numero di viaggi

Intervallo temporale dei dati per l'analisi degli incidenti

- È difficile dare un'indicazione non empirica del periodo temporale necessario per un'analisi attendibile
- La circolazione stradale risulta affetta da fattori in evoluzione, in genere non sincroni e con periodicità differenti
- In letteratura, il periodo temporale di analisi varia dai tre ai sette anni
- Per analisi *Prima-Dopo, il periodo varia da* uno o due anni prima, a uno o due anni dopo,
- Per analisi *Cross-Section,* quattro o cinque anni.

Modelli di analisi degli incidenti

- Modelli analitici che studiano la relazione tra variabili indipendenti e dati dell'incidentalità

- La relazione può riguardare
  - La presenza o no dell'incidente (binario 0/1)
  - Il numero di incidenti (conteggio)
  - La tipologia dell'incidente (multinomiale)
  - La gravità dell'incidente (ordinale)

La Regressione come modello di analisi

- I modelli sono generalmente basati sul concetto di <u>regressione</u>
- Collegano alcune variabili descrittive del guidatore, del veicolo o ambientali (indipendenti) con quella dell'incidentalità (dipendente) che può riguardare la gravità, la frequenza, o il numero
- A seconda della relazione matematica tra variabili indipendenti e dipendente si parla di:
  - Regressione lineare
  - Regressione non lineare

La regressione lineare

- Il tipico modello della regresssione lineare è

  $$\mathbf{y_i = \beta_0 + \beta_1\, x_{1i} + \ldots + \beta_p\, x_{pi}} + \varepsilon_i \qquad i = 1,\ldots, n$$

  dove

  y_i è la i-esima <u>risposta osservata del processo</u> (con funzione di distribuzione normale)

  x_p sono le <u>variabili descrittive</u> (indipendenti)

  β_p sono coefficienti da stimare

  ε_i è un termine di errore casuale ($E(\varepsilon_i=0)$)

- Il modello è di (relativamente) facile soluzione matematica ma la sua linearità si adatta poco alle non linearità presenti nei dati di incidentalità

La regressione non lineare

- Il tipico modello della regresssione non lineare è

$$\mathbf{y_i = f(x_{pi}\ ,\ \beta_{pi}) + \varepsilon_i} \qquad\qquad i = 1, \ldots, n$$

dove

y_i è la i-esima risposta osservata del processo

x_{pi} sono le variabili descrittive (indipendenti)

β_{pi} sono coefficienti da stimare e

ε_i è un termine di errore casuale ($E(\varepsilon_i=0)$)

f(.) è una <u>funzione non lineare</u>

- Molti di questi modelli possono essere convertiti in modelli lineari tramite opportuna trasformazione

Modelli lineari generalizzati (GLM)

- Per affrontare alcuni modelli non lineari sono stati proposti i modelli lineari generalizzati
- Questi modelli sono composti da una regressione lineare per η_i (dati relativi alla risposta osservata y_i)

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_p x_{pi}$$

una funzione (**link**) che lega la media di y_i, $E(y_i)=\mu$, a η

$$\mathbf{g}(\mu)=\eta$$

e una funzione che lega la varianza di y_i alla media μ

$$\mathrm{var}(y_i)=\kappa V(\mu)$$

- I dati di risposta η_i sono distribuiti in modo non normale

Esempi di Modelli GLM

Modelli tipici al variare della funzione di link, g(.), e del significato della variabile dipendente sono:

- **logistica binaria:** variabile binomiale (0/1), distribuzione di Bernoulli (link logit, probit, log-log)
- **multinomiale** estensione della regr. logistica binaria con distr. Multinomiale, e valori di risposta ordinati o non ordinati (link logit)
- **di Poisson**: valori di risposta interi positivi, distribuzione di Poisson (link log)
- **binomiale negativa**: valori di risposta interi positivi, distribuzione binomiale negativa (link log)

Modelli lineari generalizzati con effetti misti (GLMM) (1)

- I modelli lineari generalizzati con effetti misti (GLMM) consentono di includere effetti fissi o casuali sulle variabili indipendenti
- Gli effetti fissi assumono effetti costanti sui parametri, β, mentre quelli casuali possono far variare i parametri b

$$y_i = \beta_{ip} x_{pi} + b_{ip} z_{qi} + \delta_i$$

y_i è la i-esima risposta osservata del processo

x_{pi} sono le variabili legate a effetti fissi

z_{pi} sono le variabili legate a effetti casuali

β_{ip} sono coefficienti fissi da stimare

b_{ip} sono parametri variabili da stimare

δ_i è un termine di errore casuale

Modelli lineari generalizzati con effetti misti (GLMM) (2)

- Un modello GLMM ha tutti i vantaggi della regressione logistica con ulteriori vantaggi:
  - Gestisce variabili di risposta multinomiali
  - Gestisce dati non omogenei
  - Non richiede l'indipendenza dei dati (risposte y_i)
  - Dà maggiori informazioni sulla dimensione e direzione degli effetti
  - Ha una chiara struttura del modello per differenti analisi
  - Può effettuare analisi combinate con tutti gli effetti casuali in un unico modello

Modelli non lineari generalizzati (GNM)

- I modelli non lineari generalizzati (GNM) **rilassano il vincolo di linearità dei predittori**
- La maggior parte delle caratteristiche dei GLM si applica anche ai GNM
- Sono però di maggior difficoltà computazionale
- Le variabili ridondanti sono più difficili da individuare
- Appropriati valori di partenza per il calcolo (iterativo) dei parametri sono difficili da individuare
- La curva di ottimizzazione potrebbe non essere strettamente convessa
- I modelli GNM possono essere utili anche per studiare le interazioni tra variabili (indipendenti)

Modelli non lineari generalizzati con effetti misti (GNMM)

- I modelli non lineari generalizzati con effetti misti rappresentano una estensione dei modelli GNM

- Il modello GNMM è:

$$E_{y|b}(y_i)=\mu_i \quad ; \quad Var_{y|b}(y_i)=a(\phi_i)v(\mu_i)$$

$$\mathbf{g(\mu_i)=f(x_i)+z_i^T b}$$

dove

f è una funzione non lineare per gli effetti fissi

$\boldsymbol{z}_i$ sono le variabili con effetti casuali

$\boldsymbol{b}$ è un vettore di effetti casuali

I metodi di analisi *Prima-Dopo*

- Si applicano all'analisi di sezioni o intersezioni in cui siano state modificate alcune caratteristiche
- Il confronto è effettuato essenzialmente tra due gruppi di dati relativi agli incidenti prima e dopo l'intervento
- I gruppi devono essere omogenei e devono essere pesati se i periodi *Prima-Dopo* sono differenti o se i flussi sono differenti
- Il punto debole di questo approccio consiste nel numero elevato di incidenti necessari per conseguire una buona significatività
- Sono soggetti al fenomeno della "regressione alla media"

Il problema della regressione alla media "Regression to Mean"

- E' il principio che stabilisce che in misure collegate se la prima misura è o più alta o più bassa della media, il valore atteso della seconda misura è molto più vicino alla media generale
- Il grado di regressione verso la media diviene più accentuato tanto più il valore atteso della prima misura ne era lontano
- Se si prende un'area con un numero di incidenti molto superiore alla media è probabile che una successiva misura dia valori molto vicini alla media

I metodi Prima e Dopo: approcci empirici

- Particolarmente sviluppati da E. Hauer
- Il metodo naive
- Fattori correttivi
- Il gruppo di confronto
- L’approccio Bayes empirico

Tecniche Cross-Section

- Analizzano un campione di siti distinti, per un prefissato periodo, e mirano a identificare le relazioni tra gli incidenti e i parametri che li descrivono

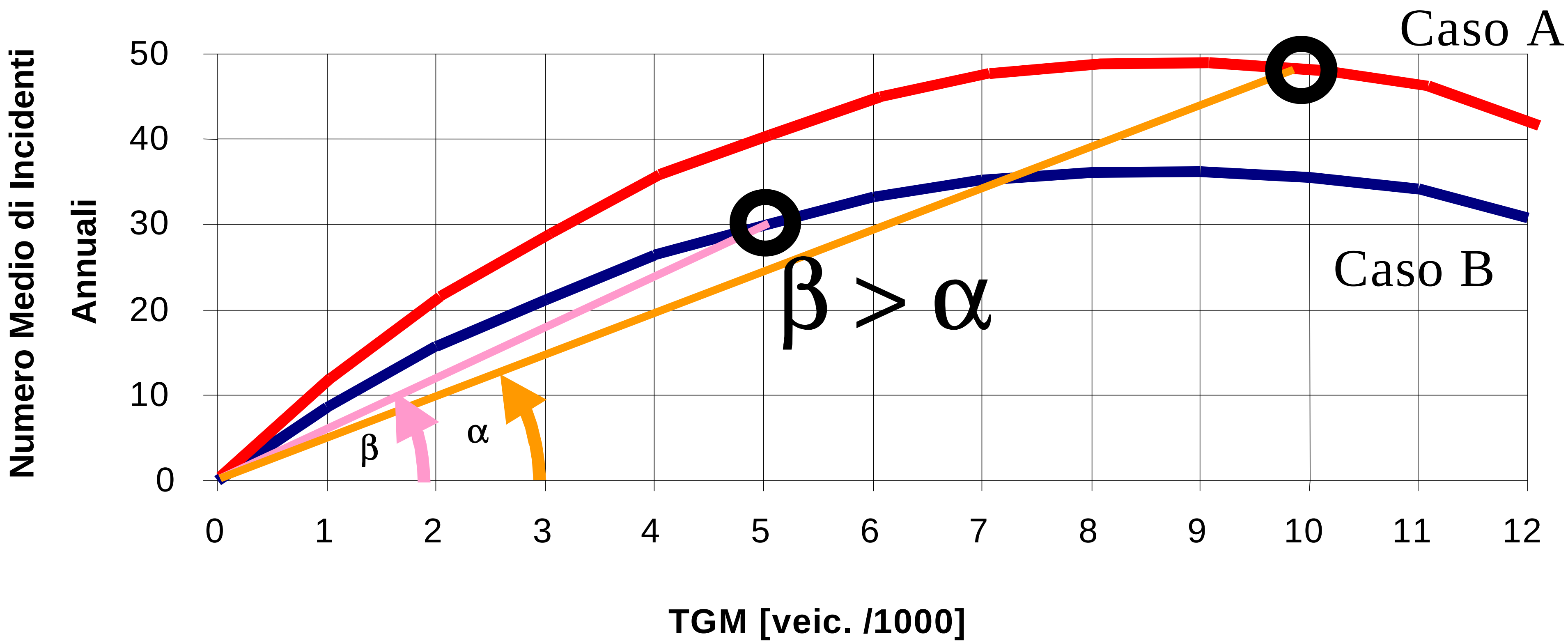
- I principali vantaggi consistono nella possibilità di realizzare una elaborazione contemporanea di tutti i parametri e nella semplicità di non dover ricorrere a due fasi di analisi temporali perché non esiste un periodo “dopo”

Problemi connessi con l'uso del tasso di incidentalità

- Se la variabile dipendente non varia linearmente con la variabile indipendente, il loro rapporto può dare risultati distorti
- Per esempio la relazione tra il flusso, q, inteso come TGM (variabile indipendente) e il numero di incidenti è per lo meno polinomiale di secondo ordine (parabola).
- Il rapporto tra il numero di incidenti e il flusso è proporzionale all'angolo formato dalla retta congiungente l'origine degli assi con il punto sulla curva
- In alcuni casi il risultato può risultare non corretto

Esempio di ambiguità nell'uso del tasso di incidentalità

Relazione Flusso-Incidenti



Grazie per l’attenzione