

Le statistiche sociali dell'ISTAT e il fenomeno degli incidenti stradali: un'esperienza di record linkage

R. Amato(), S. Bruzzone (**), V. Delmonte(***), L. Fagiolo(****)*

(*) ISTAT- Servizio Giustizia

(**) ISTAT- Servizio Sanità e assistenza

(***) ISTAT- Tirocinante (2002-2003) presso Servizio Giustizia

(****) ISTAT- Tirocinante (2002-2003) presso Servizio Sanità e assistenza

Abstract

L'incidentalità stradale è un fenomeno assai complesso, caratterizzato da aspetti che vanno da quelli più propriamente demo-sociali a quelli culturali ed economici.

I dati disponibili sul settore presentano talvolta livelli di qualità diversi e non del tutto comparabili tra loro, in quanto si riferiscono a varie rilevazioni che in alcuni casi adottano definizioni, classificazioni, nomenclature e metodologie diversi.

I bisogni informativi su tale problematica, al contrario, portano alla necessità di adottare una visione integrata del fenomeno.

INDICE

Introduzione

1. Le fonti statistiche sull'incidentalità stradale

1.1 La rilevazione sugli incidenti stradali

1.1.1 Il campo di osservazione, le unità di rilevazione e le principali definizioni

1.1.2 I modelli di rilevazione, i flussi e i soggetti coinvolti

1.2 L'indagine sulle cause di morte

1.2.1 Il campo di osservazione, le unità di rilevazione e le principali definizioni

1.2.2 I modelli di rilevazione, i flussi , i soggetti coinvolti e la codifica delle cause di morte

2. Aspetti metodologici e organizzativi

3. Studio del sottoinsieme ottenuto tramite la procedura di record linkage

4. Uno sguardo d'insieme: principali risultati

5. Analisi in componenti principali

5.1 Cenni metodologici

5.2 Principali risultati

6. Conclusioni

Bibliografia

Appendice A

Schema della metodologia applicata

Appendice B

I modelli di rilevazione

Introduzione ¹

Nell'ambito del ciclo di iniziative dell'Istat di formazione e orientamento nelle aree disciplinari e nelle specializzazioni proprie dell'Istituto, è stato avviato e concluso, per l'edizione 2002-2003, un progetto di stage finalizzato all'analisi del contributo delle statistiche sociali dell'Istat alla conoscenza del fenomeno degli incidenti stradali.

L'incidentalità stradale è un fenomeno assai complesso, caratterizzato da aspetti che vanno da quelli più propriamente demo-sociali a quelli culturali ed economici.

L'attuale base informativa sull'incidentalità stradale è caratterizzata da un panorama vasto ma molto eterogeneo poiché coinvolge fonti molteplici, alcune di natura amministrativa ed altre a carattere epidemiologico.

La diversa natura dei dati disponibili mette evidentemente in luce il problema della comparabilità delle informazioni, spesso riferite a rilevazioni che adottano definizioni, classificazioni, nomenclature e metodologie diverse. L'impostazione delle singole indagini, infatti, è spesso indirizzata ad analizzare la sicurezza stradale con un approccio mirato, mentre i bisogni informativi portano alla necessità di adottare una visione integrata dei fenomeni.

La fonte di informazione principale in tema di sicurezza sulle strade è costituita dall'indagine sugli incidenti stradali Istat-ACI, basata sulle informazioni desunte dai verbali compilati dalle autorità di polizia.

Se da un lato i rapporti di polizia restano un'insostituibile risorsa informativa sulla dinamica dell'incidente, sulle tipologie dei veicoli coinvolti, sulla localizzazione del sinistro, dall'altro sembra essere necessario affiancare a queste informazioni quelle relative all'ambiente stradale, agli esiti dell'incidente, alle lesioni riportate dagli infortunati, ai costi ospedalieri, ecc. provenienti anche da altre indagini.

Le fonti di informazioni su mobilità e incidentalità sono molte, sia all'interno dell'Istat sia presso altre Istituzioni; per quanto riguarda l'Istat un ruolo fondamentale ricopre l'indagine sulle cause di morte, ma anche l'indagine "multiscopo sulle famiglie", le statistiche giudiziarie, l'indagine sui viaggi degli italiani, l'indagine sui consumi delle famiglie, l'indagine sul trasporto merci su strada. Le fonti esterne curate da altri enti sono

La cura del presente contributo è da attribuire a Silvia Bruzzone e Lidia Fagiolo

¹ Testo di Raffaella Amato e Silvia Bruzzone (Introduzione)

rappresentate, invece, dagli archivi dell'ACI che forniscono informazioni sulla consistenza del parco veicolare, dagli archivi del Ministero dei Trasporti e delle Infrastrutture, dall'Istituto Superiore di Sanità, dagli Archivi degli Enti proprietari di strade: Stato, Regioni, Province e Comuni, dall' ANAS, dai Comuni che forniscono informazioni sulla geometria e le caratteristiche della località dove si è verificato il sinistro, dalle Strutture Sanitarie dalle quali si attingono informazioni sugli infortunati in incidenti stradali dimessi dagli ospedali, tipi di traumi, durata della degenza, costi sanitari, dagli Istituti di Assicurazione, dagli Enti gestori di strade per i dati sul traffico autostradale.

Proprio con la finalità di adottare una chiave di lettura trasversale del fenomeno dell'incidentalità stradale è stato deciso di proporre, come obiettivo di due progetti di stage, condotti con la supervisione di ricercatori-tutor, appartenenti ai Servizi Giustizia e Sanità e assistenza dell'Istat², un'analisi integrata delle indagini Istat sugli "Incidenti stradali" e sulle "Cause di morte" osservando il fenomeno da angolazioni diverse in relazione agli obiettivi conoscitivi delle indagini stesse.

Tutti i dati statistici prodotti dall'Istat, riferiti all'analisi degli incidenti stradali e delle conseguenze in termini di lesività e di mortalità, contribuiscono, se letti in una chiave unitaria, ossia, in forma integrata, ad identificare quali sono i principali fattori di rischio e, successivamente, forniscono i dati di partenza per definire una corretta politica di prevenzione.

² I progetti di stage menzionati sono stati condotti e conclusi all'interno delle due strutture: Unità Operativa GIU/C "Incidentalità stradale" del Servizio Giustizia e Unità Operativa SAN/C "Mortalità per causa" del Servizio Sanità e assistenza, le ricercatrici che hanno ricoperto la funzione di tutor sono Raffaella Amato e Silvia Bruzzone e le tirocinanti rispettivamente Valentina Delmonte e Lidia Fagiolo. Si ringraziano, inoltre, i Dirigenti dei Servizi Giustizia e Sanità ed Assistenza dell'Istat, Dott. Mario Greco e Dott.ssa Roberta Crialesi, e la Dott.ssa Luisa Frova coordinatrice delle U.O. SAN/C e SAN/D del Servizio Sanità e assistenza, per aver contribuito alla realizzazione di questo progetto.

1. Le fonti statistiche sull'incidentalità stradale ³

1.1 La rilevazione sugli incidenti stradali

1.1.1 Il campo di osservazione, le unità di rilevazione e le principali definizioni

Il campo di osservazione è costituito dall'insieme degli incidenti stradali avvenuti sul territorio nazionale nell'arco di un anno solare.

L'unità di rilevazione è il singolo incidente stradale.

La rilevazione è riferita al momento in cui l'incidente si è verificato. E' con riguardo a tale momento, quindi, che di ciascuna unità vengono considerati i caratteri e le modalità, le cause o le circostanze determinanti, le conseguenze sia per le persone che per le cose.

Di seguito si riportano le principali definizioni utilizzate nell'indagine:

- **incidenti stradali:** risultano quelli che si verificano in una strada aperta alla circolazione pubblica, in seguito ai quali una o più persone sono rimaste ferite o uccise e nei quali almeno un veicolo è rimasto implicato.

Prima del 1991 l'Istat rilevava tutti gli incidenti stradali, anche quelli che non necessariamente comportavano lesioni alle persone ma solo danno alle cose. La definizione attuale di incidente stradale dà luogo ad un concetto più interessante sotto il profilo dell'analisi, poiché restringendo il campo di osservazione ai soli incidenti che causano danno alle persone si ottiene una lettura più corretta e mirata dei sinistri più gravi; inoltre, permette di effettuare confronti internazionali.

- **morti:** le persone decedute sul colpo (entro le 24 ore) o quelle decedute dal 2° al trentesimo giorno, a partire da quello dell'incidente compreso.

Tale definizione, anch'essa conforme alle norme internazionali, si applica agli incidenti stradali verificatisi a partire dal 1° gennaio 1999. Prima di tale data il periodo di tempo necessario per determinare il numero dei decessi era pari a sette giorni dal momento dell'incidente.

- **feriti:** le persone che hanno subito lesioni al proprio corpo a seguito dell'incidente.

Data la difficoltà di definire criteri obiettivi sul livello di gravità delle lesioni subite, non si distingue tra feriti gravi o leggeri.

³ Testo di Raffaella Amato (par. 1.1) e Silvia Bruzzone (par. 1.2)

1.1.2 I modelli di rilevazione, i flussi e i soggetti coinvolti

La rilevazione, come accennato già sopra, viene effettuata mediante autocompilazione del questionario cartaceo o informatizzato (modello Istat/CTT/INC).

Gli organi rilevatori (Polizia Stradale, Carabinieri e Polizia Municipale), a seconda del proprio grado di informatizzazione, trasmettono i dati elementari all'Istat inviando i modelli compilati o i file già registrati su supporto magnetico (floppy disk o e-mail).

La qualità del dato statistico prodotto è fortemente legata al livello di collaborazione dei rispondenti.

La rilevazione è eseguita con la collaborazione della Polizia Stradale, i Carabinieri e la Polizia Municipale preposti al controllo della circolazione e alla disciplina del traffico.

In particolare, la rilevazione viene effettuata tramite la compilazione del “rapporto statistico d'incidente stradale” (modello di rilevazione) da parte dell'autorità che è intervenuta sul luogo del sinistro in cui è coinvolto almeno un veicolo in circolazione sulla rete viaria. Tutti i modelli di rilevazione contenenti i dati elementari relativi agli incidenti stradali vengono inviati all'Istat che provvede al loro controllo, all'elaborazione, all'analisi e alla diffusione delle statistiche prodotte.

I principali caratteri rilevati sono:

- Data e Località dell'incidente
- Organo di rilevazione
- Localizzazione dell'incidente: fuori dalla zona abitata o nell'abitato
- Tipo di strada
- Pavimentazione
- Fondo stradale
- Segnaletica
- Condizioni meteorologiche
- Natura dell'incidente (scontro, fuoriscita, investimento.....)
- Tipo di veicoli coinvolti
- Circostanze dell'incidente
- Conseguenze dell'incidente alle persone e ai veicoli

1.2 L'indagine sulle cause di morte

1.2.1 Il campo di osservazione, le unità di rilevazione e le principali definizioni

Il campo di osservazione dell'indagine è costituito dall'insieme dei decessi verificatisi, sul territorio nazionale, nell'arco di un determinato periodo di riferimento (anno solare).

L'unità di rilevazione ed analisi è rappresentata dal singolo individuo deceduto.

Accanto ai codici della Classificazione Internazionale delle malattie rivestono un ruolo molto importante anche le definizioni delle entità che entrano in gioco sia nella fase di certificazione che nella fase di codifica. Riportiamo qui di seguito alcune importanti definizioni utilizzate nell'indagine nel caso di decesso per incidente stradale :

- **incidente stradale da veicolo a motore:** è ogni accidente da veicolo a motore che si verifica sulla pubblica via (ad es. che origina, termina o coinvolge un veicolo parzialmente sulla pubblica via). Si presume che un accidente da veicolo a motore abbia luogo sulla pubblica via, a meno che non sia specificato un posto diverso. Fanno eccezione i casi di incidenti che coinvolgono solamente i veicoli a motore “fuori strada”, classificati come incidenti non stradali, a meno che non sia dichiarato il contrario.
- **via pubblica (via di traffico o strada):** è l'intera estensione di ogni via o piazza compresa fra i limiti di proprietà (o di altre linee di demarcazione), di cui una parte sia aperta al pubblico per la circolazione dei veicoli sia per il diritto che per consuetudine.
- **carreggiata:** è quella parte della via pubblica progettata, costruita e ordinariamente utilizzata per la circolazione dei veicoli.
- **veicolo a motore:** è ogni mezzo mosso meccanicamente o elettricamente, non circolante su rotaie, su cui possono essere trasportate persone o cose sulla pubblica via. Ogni oggetto trainato da un veicolo a motore, come un rimorchio, un carrello, una slitta, un vagone è considerato parte del veicolo a motore.

1.2.2 I modelli di rilevazione, i flussi , i soggetti coinvolti e la codifica delle cause di morte

I modelli di rilevazione

L'indagine sulle cause di morte" viene effettuata correntemente dall'ISTAT attraverso l'utilizzo dei modelli ISTAT/D.4 (scheda di morte per maschio oltre il primo anno di vita), ISTAT/D.5 (scheda di morte per femmina oltre il primo anno di vita), ISTAT/D.4 bis (scheda di morte per maschio nel primo anno di vita), ISTAT/D.5 bis (scheda di morte per femmina nel primo anno di vita). Su tali modelli vengono riportate le notizie relative al decesso fornite dal medico curante o necroscopo (Parte A della scheda di morte) e le informazioni di carattere demografico e sociale (Parte B della scheda di morte) a cura dell'ufficiale di Stato Civile.

Le informazioni richieste nella Parte A della scheda di morte, a cura del medico sono:

- provincia e comune di decesso;
- descrizione della lesione;
- malattie o complicazioni eventualmente sopravvenute a seguito della lesione;
- stati morbosi preesistenti che hanno eventualmente contribuito al decesso;
- mezzo o modo col quale la lesione è stata determinata
- data dell'accidente (ora, gg/mm/aaaa)
- intervallo tra l'azione violenta e la morte
- luogo dell'accidente.

Le principali variabili di natura demografica e sociale (Parte B) rilevate sono:

- località dell'evento - provincia e comune di decesso;
- data di morte (giorno, mese, anno);
- data di nascita (giorno, mese, anno);
- località di nascita - comune e provincia o stato estero;
- età compiuta (in anni se si tratta di un individuo oltre il 1° anno di vita, in ore, giorni o mesi se si tratta di un decesso nel 1° anno di vita);
- stato civile (se coniugato viene richiesto di specificare l'anno di nascita del coniuge superstite e l'anno di matrimonio);
- località di residenza - comune e provincia o stato estero;
- grado di istruzione (nel caso di individui deceduti nel 1° anno di vita, l'informazione si riferisce alla condizione dei genitori);

- condizione professionale o non professionale, posizione nella professione e ramo di attività economica (nel caso di individui deceduti nel 1° anno di vita, l'informazione si riferisce alla condizione dei genitori);
- cittadinanza (se straniera viene richiesto di indicare il codice cittadinanza straniera);
- codice individuale del deceduto.

I flussi

Ogni anno, l'Istat provvede ad inviare le nuove edizioni dei modelli D4, D5, D4bis, D5bis agli Uffici Regionali e a tutti i comuni italiani (8100). Le aziende sanitarie locali (ASL), gli ospedali e i medici di famiglia devono rivolgersi, infatti, agli Uffici Comunali di competenza o gli Uffici Regionali per effettuare il ritiro delle schede cartacee da compilare.

Dopo avere completato la parte sanitaria, il medico provvede ad inviare il modello ai comuni.

L'ufficiale di stato civile inserisce, nella parte A del modello, le informazioni demografiche del deceduto.

Copia della scheda di morte (copia per la ASL) deve essere inviata entro trenta giorni, dal comune ove è avvenuto il decesso all'unità sanitaria locale nel cui territorio detto Comune è ricompreso. Qualora il deceduto fosse residente nel territorio di una unità sanitaria locale diversa da quella ove è avvenuto il decesso, quest'ultima deve inviare copia della scheda di morte all'unità sanitaria locale di residenza. Nel caso di comuni comprendenti più unità sanitarie locali, tali comunicazioni sono dirette a quella competente. (Cfr. Art. 7 - Regolamento di Polizia Mortuaria D.P.R. n. 285 del 10 settembre 1990). Una seconda copia viene inviata alla sede centrale dell'Istat di Roma, dopo essere passata per le Prefetture e agli Uffici Istat regionali, i quali sono tenuti ad effettuare un primo controllo quantitativo del materiale pervenuto.

I soggetti coinvolti

Il medico certificatore è tenuto a compilare tutti i quesiti (Parte A dei modelli D.4, D.5, D.4bis, D.5bis) relativi alla causa di decesso presenti sulla scheda di morte ed in particolare deve riportare la “causa iniziale” ossia la malattia che attraverso eventuali complicazioni o stati morbosi intermedi ha condotto al decesso, la “causa intermedia o complicazione” che include l’eventuale successione morbosa della malattia indicata al primo quesito, la “causa terminale” per la quale si intende la malattia o lo stato morboso che ha direttamente provocato il decesso e gli “altri stati morbosi rilevanti” tra i quali si considerano le malattie e gli stati morbosi che hanno contribuito al decesso. Nel caso in cui si tratti di causa di morte violenta il medico dovrà anche indicare con esattezza la “descrizione della lesione”, “le malattie o complicazioni”, “gli stati morbosi preesistenti” ed il “mezzo o modo” col quale la lesione è stata determinata.

La compilazione della scheda termina con la data, la firma leggibile, sotto la dichiarazione che le notizie date in “scienza e coscienza” sono vere e con l’indicazione della qualifica, se curante o necroscopo, del medico certificatore.

La Parte B della scheda di morte contiene le informazioni a cura dell’Ufficiale di Stato Civile riguardanti le notizie di natura demografica, sociale ed economica.

Accanto alle informazioni di tipo sanitario, riferite al processo morboso, il medico certificatore è tenuto ad indicare anche la data dell’accidente (ora, gg/mm/aaaa), l’intervallo tra l’azione violenta e la morte, il luogo dell’accidente e l’intervallo di tempo intercorso tra la lesione e la morte.

La codifica delle cause di morte

L’identificazione della causa iniziale ed il mezzo o modo di lesione viene effettuata sulla base delle raccomandazioni dell’Organizzazione Mondiale della Sanità. Le regole su cui si basa la codifica sono fondate sulla nozione di causa iniziale, cioè lo stato morboso o il traumatismo che è all’origine del concatenamento che ha prodotto il decesso. Quando la causa è un traumatismo, la codifica è fatta tenendo conto delle cause esterne dello stesso o della sua natura o di entrambi.

La classificazione supplementare delle cause esterne dei traumatismi ed avvelenamenti (codici E ICD9 Rev.1975) viene utilizzata per la codifica delle circostanze e delle condizioni ambientali dei traumatismi stessi. Ogni qual volta si utilizza un codice della classificazione supplementare si intende che esso sia adoperato in aggiunta ad un codice del settore del Settore della ICD che indica la natura della causa morbosa.

2. Aspetti metodologici ed organizzativi ⁴

L'obiettivo principale del presente lavoro, come già esposto precedentemente, è di verificare l'effettiva possibilità di integrazione delle informazioni desunte dalle due diverse fonti sull'incidentalità stradale in Italia: “Incidenti stradali” e “Cause di morte”.

A tal fine è stata utilizzata una procedura di *record linkage*.

Tale procedura è una tecnica algoritmica che ha lo scopo di identificare quali coppie di records di due basi di dati corrispondono ad una stessa unità.

I metodi per il record linkage si possono dividere in due grandi gruppi: metodi *ad hoc* o euristici e metodi basati su modello, o “statistici”.⁵

Nonostante siano molti i motivi che portano a preferire i metodi statistici in questo ambito si fa riferimento ad una procedura del tutto deterministica (per maggiori dettagli si rimanda all'appendice A).

Di seguito si elencano i passi preliminari compiuti per poter applicare la procedura di record linkage messa a punto.

Per poter collegare le informazioni derivanti dalle due indagini Istat è stato necessario identificare una *chiave di linkage*, ossia un set di variabili comuni alle due basi di dati che ne costituisce il matching.

Per mettere a punto una chiave di linkage che conducesse ad una buona performance della procedura di associazione, si è deciso di considerare per l'indagine sulle cause di morte, i dati provenienti direttamente dalla fase di registrazione, che contengono un maggior numero di informazioni rispetto al file standard diffuso agli utenti.

In questo modo si sono recuperate alcune variabili (ora, giorno, mese e anno dell'incidente e giorno del decesso) presenti nella rilevazione degli incidenti e molto rilevanti per la definizione della chiave.

Una conseguenza di questa scelta è stata l'esclusione dall'analisi dei record relativi alle province autonome di Trento e Bolzano.

⁴ Testo di Valentina Delmonte

⁵ Cfr. Istat, Metodi e Norme n.16 (2003), *Metodi statistici per il record linkage*, Roma.

Queste ultime, infatti, grazie ad una convenzione stipulata con l'Istat a metà degli anni '90, inviano i dati relativi all'indagine già codificati, per quanto riguarda le cause e, direttamente su supporto informatico.

I dati acquisiti pur contenendo tutte le informazioni necessarie per la successiva diffusione agli utenti, non riportano tutte le notizie presenti nella scheda di morte.

Per rendere omogeneo dal punto di vista dei contenuti informativi il file dei dati sulle cause di morte e successivamente le due popolazioni oggetto di studio si è deciso di non includere nell'analisi i dati di Trento e Bolzano.

Con la scheda di morte vengono rilevati comune e provincia di decesso mentre nel Rapporto statistico d'incidente stradale comune e provincia di sinistro.

Per risolvere tale problema e per aumentare lo spazio delle variabili comuni alle due popolazioni, si è ritenuto opportuno indicare, nell'indagine sulle cause di morte, comune e provincia di decesso come luoghi d'incidente.

Questa soluzione ha permesso di ottenere risultati più affidabili ma più contenuti in quanto non sempre il luogo dell'incidente corrisponde a quello del decesso.

La chiave di linkage utilizzata per associare i dati individuali è stata ottenuta tramite la combinazione delle seguenti variabili comuni:

- Mese dell'incidente
- Giorno dell'incidente
- Provincia di incidente
- Comune di incidente
- Sesso
- Età

Dopo un'attenta analisi sono stati eliminati i records non informativi per i quali, cioè, non erano indicati i valori delle suddette variabili.

Mentre nel modello relativo agli incidenti stradali si considerano i decessi occorsi entro 30 giorni dall'incidente nell'anno di rilevazione, nell'indagine sulle cause di morte sono rilevati tutti i decessi avvenuti nell'anno, indipendentemente dall'intervallo tra data

incidente e data morte. Alla luce di tale considerazione sono stati eliminati dal file sulle cause di morte i decessi avvenuti oltre i 30 giorni.

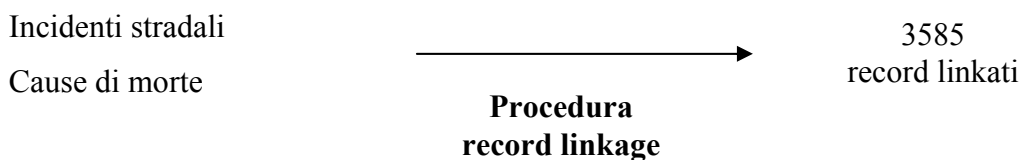
Di seguito è fornito un breve schema che illustra le fasi di questo studio preliminare dei dati:

- Eliminazione dei dati non informativi (in entrambi i file)
- Eliminazione dei dati con intervallo tra data accidente e data morte superiore ai 30 giorni (nel file sulle cause di morte)
- Eliminazione dei dati relativi a Trento e Bolzano per incompatibilità di informazioni con le altre province (in entrambi i file)



<u>Numerosità iniziale popolazione</u>	→	<u>Numerosità finale popolazione</u>
6410	Incidenti stradali	6002
7003	Cause di morte	6126

Attraverso la procedura di linkage è stato possibile associare 3585 record, che rappresentano circa il 60% e il 59 % rispettivamente della popolazione di riferimento degli incidenti stradali e delle Cause di morte.



3. Studio del sottoinsieme ottenuto attraverso la procedura di record linkage⁶

In questa parte del lavoro sono state analizzate le distribuzioni di alcune variabili per valutare se quelle relative al sottoinsieme ottenuto attraverso la procedura di linkage risultassero simili a quelle delle popolazioni di origine.

Le tabelle che seguono mostrano la distribuzione per sesso e i valori degli indici di posizione e dei momenti della variabile età per i tre collettivi oggetto di studio:

Tabella 1 - Distribuzione per sesso.

Sesso	Cause di morte		Sottoinsieme		Incidenti stradali	
	N	Percentuale	N	Percentuale	N	Percentuale
Maschi	4650	75,9	2760	77,0	4634	77,2
Femmine	1476	24,1	825	23,0	1368	22,8
Totale	6126	100,0	3585	100,0	6002	100,0

Tabella 2 - Indici di posizione e momenti della variabile età.

	Maschi			Femmine		
	Sottoinsieme	Cause di morte	Incidenti stradali	Sottoinsieme	Cause di morte	Incidenti stradali
Media	41.89	42.59	42.54	46.12	47.78	46.02
Mediana	36	36	36	44	48	43
Moda	25	25	25	23	20	22
Asimmetria	0.61	0.54	0.54	0.17	0.06	0.19
Curtosi	0.61	0.54	0.54	-1.26	-1.32	-1.24

I risultati riportati nelle tabelle vengono meglio evidenziati dai grafici a foglia seguenti che rappresentano le distribuzioni per sesso della variabile età per i tre collettivi in esame.

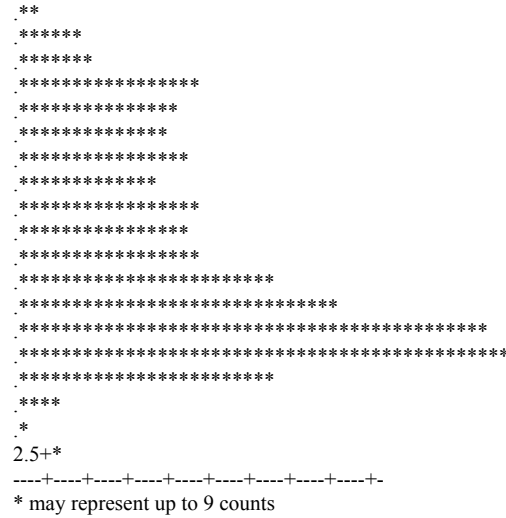
⁶ Testo di Valentina Delmonte

Variabile= ETA

SESSO=M

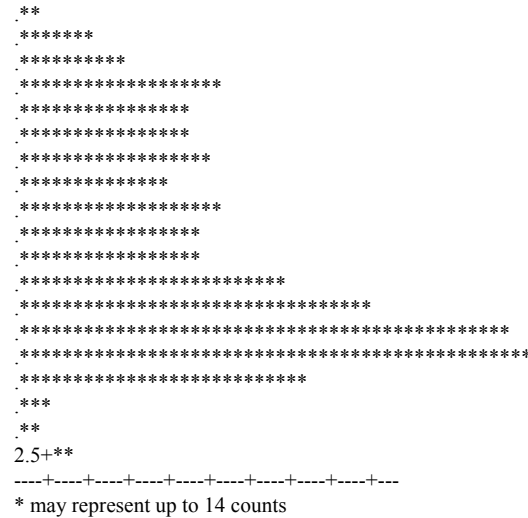
Histogram
97.5+*

Sottoinsieme



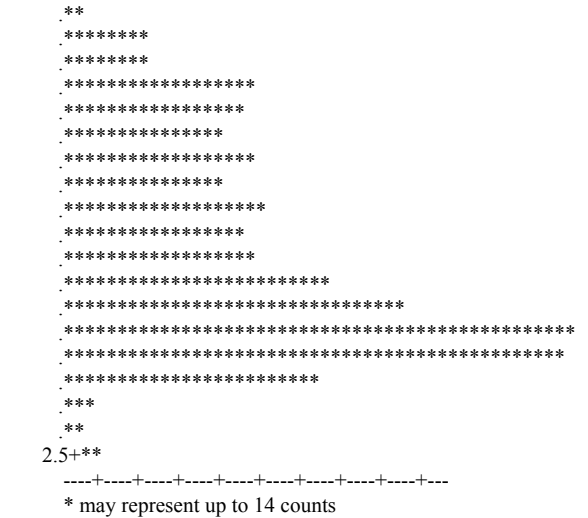
Histogram
97.5+*

Cause morte



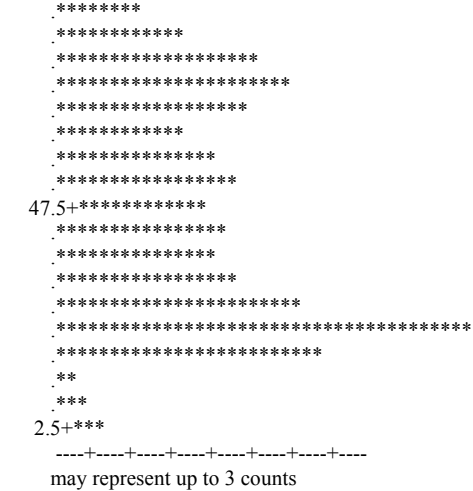
Histogram
97.5+*

Incidenti stradali



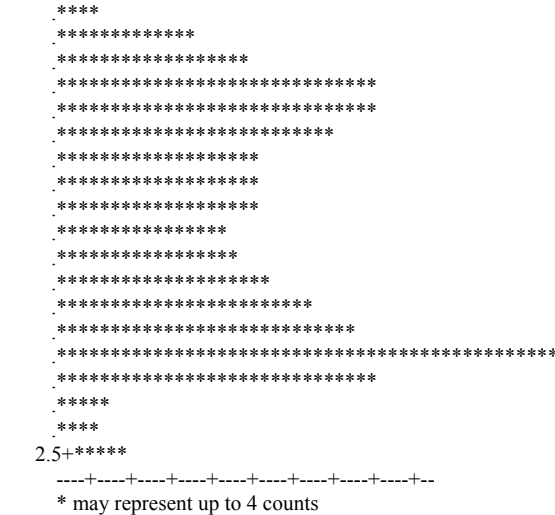
Histogram
92.5+***

SESSO=F



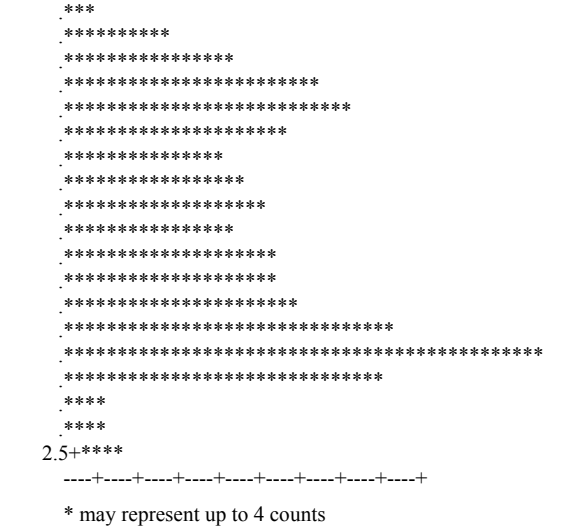
Histogram
97.5+*

Cause morte



Histogram
97.5+*

Incidenti stradali



Successivamente la variabile età è stata ricodificata in classi quinquennali e la variabile provincia di decesso in regioni di decesso (Cfr tabelle 3-4 e grafici 1-2)

Tabella 3 - Distribuzione per classi d'età.

Classi d'età	Cause di morte		Sottoinsieme		Incidenti stradali	
	N	Percentuale	N	Percentuale	N	Percentuale
1-4	14	0,39	35	0,57	29	0,48
5-9	14	0,39	34	0,56	35	0,58
10-14	33	0,92	59	0,96	55	0,92
15-19	288	8,03	494	8,06	455	7,58
20-24	530	14,78	845	13,79	830	13,83
25-29	458	12,78	750	12,24	789	13,15
30-34	318	8,87	542	8,85	526	8,76
35-39	261	7,28	417	6,81	426	7,10
40-44	196	5,47	299	4,88	331	5,51
45-49	175	4,88	295	4,82	297	4,95
50-54	204	5,69	333	5,44	342	5,70
55-59	158	4,41	270	4,41	271	4,52
60-64	174	4,85	318	5,19	299	4,98
65-69	172	4,80	323	5,27	293	4,88
70-74	191	5,33	337	5,50	333	5,55
75-79	204	5,69	376	6,14	339	5,65
80-84	97	2,71	208	3,40	175	2,92
85+	98	2,73	191	3,12	177	2,95
Totale	3585	100,0	6126	100,0	6002	100,0

Grafico 1 - Distribuzione per classi di età.

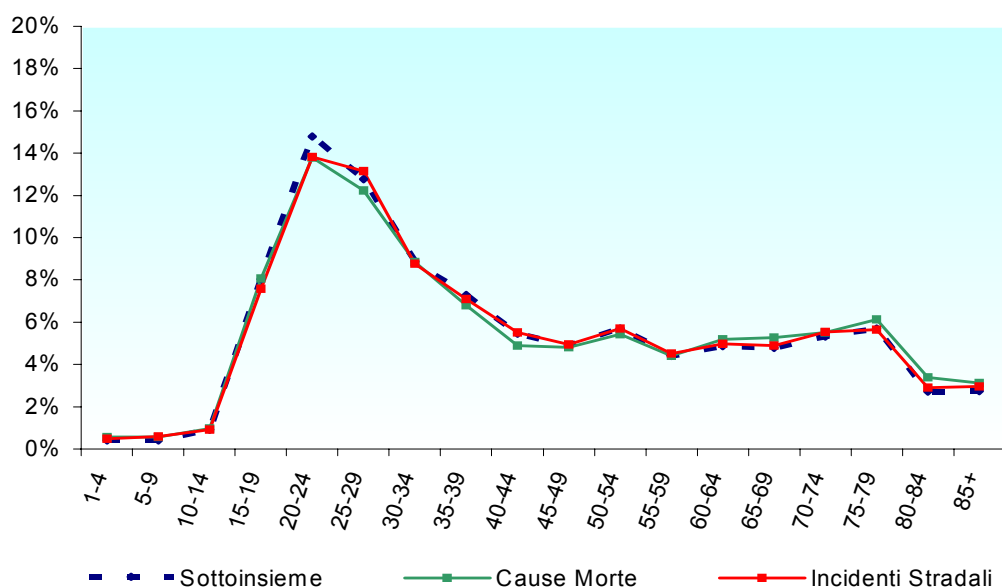
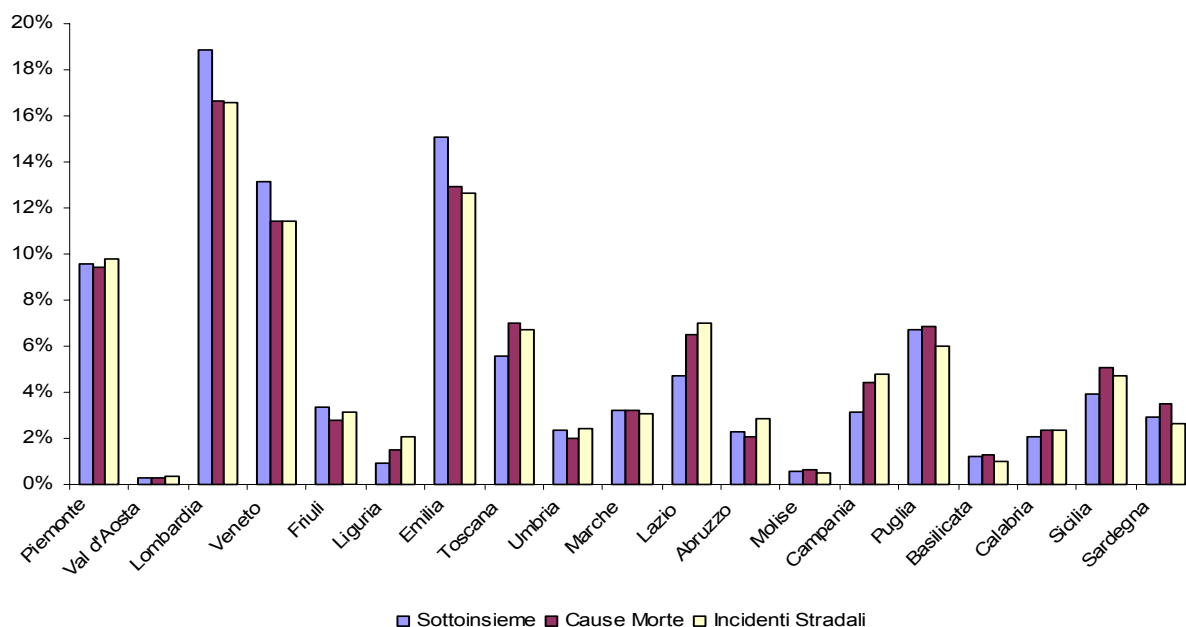


Tabella 4 - Distribuzione per regione di decesso.

	Cause di morte		Sottoinsieme		Incidenti stradali	
	N	Percentuale	N	Percentuale	N	Percentuale
Piemonte	342	9,54	578	9,44	586	9,76
Val d'Aosta	11	0,31	18	0,29	21	0,35
Lombardia	677	18,88	1020	16,65	996	16,59
Veneto	472	13,17	700	11,43	686	11,43
Friuli	120	3,35	170	2,78	189	3,15
Liguria	34	0,95	94	1,53	123	2,05
Emilia	541	15,09	790	12,90	758	12,63
Toscana	200	5,58	429	7,00	404	6,73
Umbria	85	2,37	123	2,01	145	2,42
Marche	115	3,21	197	3,22	184	3,07
Lazio	170	4,74	398	6,50	418	6,96
Abruzzo	81	2,26	129	2,11	173	2,88
Molise	21	0,59	39	0,64	32	0,53
Campania	113	3,15	270	4,41	286	4,77
Puglia	240	6,69	422	6,89	358	5,96
Basilica	44	1,23	80	1,31	59	0,98
Calabria	73	2,04	144	2,35	142	2,37
Sicilia	140	3,91	312	5,09	285	4,75
Sardegna	106	2,96	213	3,48	157	2,62
Totale	3585	100,00	6126	100,00	6002	100,00

Grafico 2 - Distribuzione per regione di decesso.



I due grafici mostrati confermano che il sottoinsieme ottenuto ha caratteristiche molto simili alle due popolazioni di partenza anche in relazione alle due variabili ricodificate.

4. Uno sguardo d'insieme: principali risultati ⁷

Il lavoro svolto, come già esposto nelle pagine precedenti, rappresenta un primo tentativo di integrazione delle informazioni relative agli incidenti stradali desunte dalle due indagini Istat su “Incidenti stradali” e “Cause di morte”; proprio per questo motivo si è preferito dare risalto allo studio congiunto di alcune variabili provenienti dalle due diverse fonti e mai analizzate fino ad ora, invece di soffermarsi sulle analisi “classiche” di cui si può prendere visione dai volumi pubblicati dall'Istat.

Si ritiene importante ribadire che i risultati che vengono presentati sono relativi ad un collettivo costituito dai solo records per i quali la procedura di record linkage ha avuto successo e quindi non ci consentono di effettuare generalizzazioni. La loro importanza sta nel fatto che permettono, per la prima volta, di analizzare congiuntamente alcuni dei diversi aspetti della sinistrosità.

Negli ultimi anni in Italia, come in tutti i paesi a motorizzazione matura, si assiste ad un continua riduzione del numero di morti per incidenti stradali (in termini assoluti dal 1990 al 2000 si è passati da 9123 a 7262 vittime).

Gli incidenti tendono, dunque, a diventare “meno gravi” e tra i diversi fattori che hanno reso possibile questo risultato (quali i miglioramenti introdotti nella sicurezza attiva e passiva dei veicoli, la maggiore efficienza delle strutture sanitarie..) un ruolo importante rivestono le nuove normative in materia di sicurezza che hanno reso obbligatorio l'uso del casco e delle cinture.

E' ormai assodato, sulla base di un'ampia letteratura epidemiologica internazionale, che l'adozione dei dispositivi di sicurezza riduce il rischio di morte o di lesioni gravi intorno al 50%.⁸

Alla luce di queste considerazioni si è ritenuto interessante, nonostante il sottoinsieme oggetto di studio sia costituito da individui rimasti vittime di incidenti, analizzare l'uso delle protezioni obbligatorie in relazione ad alcune variabili socio-demografiche (come detto in precedenza l'uso di tali dispositivi evita nella maggioranza dei casi che gli incidenti abbiano conseguenze gravi ma purtroppo non garantisce l'incolumità).

⁷ Testo di Lidia Fagiolo.

⁸ Cfr. Taggi F. (a cura di), (2003) *Aspetti sanitari della sicurezza stradale* (secondo rapporto sul progetto DATIS). Istituto Superiore di Sanità, Roma

Tra gli individui del sottoinsieme la percentuale di utilizzo delle protezioni risulta molto elevata per entrambi i sessi: il 57,6% per i maschi e il 65,2% per le femmine.

Tabella 5 - Uso delle protezioni obbligatorie per sesso.

Uso delle protezioni obbligatorie	Maschi		Femmine	
	N	Percentuale	N	Percentuale
Si	1590	57,6	538	65,2
No	584	21,2	92	11,2
Non indicato	586	21,2	195	23,6
Totale	2760	100,0	825	100,0

Per quanto riguarda l'uso dei dispositivi di sicurezza in relazione con l'età si può notare che la percentuale più bassa di utilizzo si ha, sia per i maschi che per le femmine del collettivo, nella fascia di età "55 e oltre". Tra coloro che sono rimasti vittime di incidenti stradali, infatti, il 32,2% e il 19,3%, rispettivamente per i maschi e le femmine, di ultracinquantenni non utilizza le protezioni obbligatorie, contro percentuali che si aggirano intorno al 25 e 12, rispettivamente per maschi e femmine, nelle altre classi di età. (Cfr. tabella 6)

Tabella 6 - Uso delle protezioni obbligatorie per sesso ed età.

Classe di età	N			Percentuale		
	Si	No	Totale	Si	No	Totale
	Maschi					
<30	634	204	838	75,7	24,3	100,0
30-54	570	197	767	74,3	25,7	100,0
55 e oltre	386	183	569	67,8	32,2	100,0
	Femmine					
<30	188	24	212	88,7	11,3	100,0
30-54	158	22	180	87,8	12,2	100,0
55 e oltre	192	46	238	80,7	19,3	100,0

Sembra interessante sottolineare che la percentuale più alta di utilizzatori è tra i celibi e le nubili (rispettivamente 75,7% e 87,9%)

Tabella 7 - Uso delle protezioni obbligatorie per sesso e stato civile. (a) (b)

Stato civile	N			Percentuale		
	Si	No	Totale	Si	No	Totale
	Maschi					
Celibi	843	271	1114	75,7	24,3	100,0
Coniugati	625	265	890	70,2	29,8	100,0
Vedovi	63	24	87	72,4	27,6	100,0
Divorziati	30	12	42	71,4	28,6	100,0
	Femmine					
Nubili	203	28	231	87,9	12,1	100,0
Coniugate	226	39	265	85,3	14,7	100,0
Vedove	94	21	115	81,7	18,3	100,0
Divorziate	10	3	13	76,9	23,1	100,0

(a) I separati sono stati associati ai coniugati

(b) Per 60 individui, pari al 1,7% del totale, non è indicato lo stato civile

Infine è stato analizzato l'uso delle protezioni obbligatorie in relazione al titolo di studio posseduto: per i maschi del sottoinsieme sono maggiormente i laureati a farne uso, per le donne la percentuale più alta è relativa a coloro che hanno la media inferiore. (Cfr. tabella 8)

Tabella 8 - Uso delle protezioni obbligatorie per sesso e titolo di studio. (c)

Stato civile	N			Percentuale		
	Si	No	Totale	Si	No	Totale
Maschi						
Laurea	70	13	83	84,3	15,7	100,0
Media superiore	295	78	373	79,1	20,9	100,0
Media inferiore	635	219	854	74,4	25,6	100,0
Licenza elementare	413	203	616	67,0	33,0	100,0
Femmine						
Laurea	23	3	26	88,5	11,5	100,0
Media superiore	102	18	120	85,0	15,0	100,0
Media inferiore	182	15	197	92,4	7,6	100,0
Licenza elementare	177	43	220	80,5	19,5	100,0

(c) Per 387 individui, pari all' 11% del totale, non è indicato il titolo di studio

Passando ora all'analisi delle informazioni provenienti dalle Indagine sulle cause di morte concentriamo ora l'attenzione sulla natura delle lesioni che hanno provocato il decesso delle persone coinvolte negli incidenti stradali e sulle modalità che hanno provocato le suddette lesioni.

Prima di presentare i risultati si ritiene opportuno spendere qualche parola sulle informazioni sanitarie fornite dalla rilevazione sulle cause di morte.

Le statistiche di mortalità per causa, come già detto⁹, sono basate convenzionalmente su una singola causa di morte, la cosiddetta "causa iniziale" la cui definizione è sancita dall'Organizzazione Mondiale della Sanità (OMS) come segue:

- 1) "la malattia o il traumatismo che avvia il concatenamento degli eventi morbosi che conduce direttamente alla morte", o
- 2) "l'insieme delle circostanze dell'accidente o della violenza che hanno provocato la lesione traumatica mortale".

L'individuazione e la codifica della causa primaria avviene quindi sulla base di opportuni criteri di decisione, in accordo con le regole di codifica fornite dall'OMS, tenendo conto di tutte le informazioni demografiche e sanitarie riportate sulla scheda di morte.

Nel caso si tratti di una causa di morte violenta ed in particolare di un incidente stradale, accanto al codice causa iniziale, compreso nel settore XVII della Classificazione Internazionale delle

⁹ Cfr. Cap.2 par.2.2.2

Malattie ICD IX Revisione 1975 ¹⁰ “Traumatismi ed avvelenamenti” (codici ICD IX 800-999), viene aggiunto un secondo codice, riferito al mezzo o modo con cui la lesione è stata causata, appartenente alla classificazione E della ICD IX e compreso nel gruppo E810-E819 riferito agli incidenti stradali da veicolo a motore.

Per i codici di questa categoria è prevista l’aggiunta di sottocategorie, rappresentate dalla quarta cifra dei codici ICD IX del gruppo E810-E819, utilizzate per specificare la persona coinvolta e deceduta nell’incidente, come ad esempio conducente, passeggero, motociclista, pedone etc..

Sono i traumatismi interni dell’addome del bacino e di sede non specificata e le fratture del cranio e della faccia le principali cause di morte per le vittime del sottoinsieme oggetto di studio; il 55,8% degli incidenti stradali risulta di natura non specificata (Cfr. tabelle 9-10).

Tabella 9 - Natura delle lesioni che hanno portato alla morte.

	N	Percentuale
Traumatismi interni dell’ addome, del bacino e di sede non specificata	1.258	35,1
Fratture del cranio e della faccia	1.195	33,3
Traumatismi intracranici (esclusi quelli associati a fratture del cranio e della faccia	419	11,7
Traumatismi interni del torace	335	9,3
Fratture del collo, del tronco e degli arti superiori	158	4,4
Fratture del femore e degli arti inferiori	114	3,2
Altro	106	3,0
Totale	3585	100,0

Tabella 10 - Modalità che hanno provocato la lesione.

	N	Percentuale
Accidente stradale da veicolo a motore di natura non specificata.	2.002	55,8
Accidente stradale da veicolo a motore con scontro con altro veicolo.	884	24,7
Accidente stradale da veicolo a motore con scontro con pedone.	358	10,0
Accidente stradale da veicolo a motore senza scontro, dovuto a perdita del suo controllo.	341	9,5
Totale	3.585	100,0

Infine si è ritenuto interessante focalizzare l’attenzione sull’ora in cui è avvenuto l’incidente nel quale hanno perso la vita gli individui del collettivo considerato.

Il numero di incidenti verificatesi di notte (tra lo ore 22 e le 6 del mattino) è particolarmente elevato per i celibi e le nubili e questi ultimi hanno età inferiore ai trent’anni. (Cfr. tabelle 11-12)

¹⁰ Cfr. Istat, Metodi e Norme (1998), *Classificazione delle malattie, traumatismi e cause di morte. 9ª revisione 1975.*

Questo dato sembra evidenziare un diverso uso dell'automobile da parte dei giovani: più per svago che per lavoro.

Tabella 11 - Decessi per sesso, stato civile e ora dell'incidente. (a) (b) (d)

	Notte		Giorno	
	N	Percentuale	N	Percentuale
Maschi				
Celibi	634	72,9	744	41,2
Coniugati	205	23,6	928	51,4
Vedovi	13	1,5	102	5,7
Divorziati	18	2,1	31	1,7
Totale	870	100,0	1805	100,0
Femmine				
Nubili	131	62,1	173	29,0
Coniugate	60	28,4	273	45,7
Vedove	13	6,2	142	23,8
Divorziate	7	3,3	9	1,5
Totale	211	100,0	597	100,0

(a) I separati sono stati associati ai coniugati

(b) per 60 individui, pari al 1,7% del totale, non è indicato lo stato civile

(d) per 43 individui, pari al 1,2% del totale, non è indicata l'ora dell'incidente

Tabella 12 - Decessi per sesso, stato civile, età e ora dell'incidente. (a)

	Notte				Giorno			
	<30	30-54	55 e oltre	Totale	<30	30-54	55 e oltre	Totale
Maschi								
Celibi	75,1	23,0	1,9	100,0	65,1	25,7	9,3	100,0
Coniugati	8,3	62,9	28,8	100,0	4,1	41,6	54,3	100,0
Vedovi	0,0	15,4	84,6	100,0	0,0	6,9	93,1	100,0
Divorziati	0,0	83,3	16,7	100,0	0,0	61,3	38,7	100,0
Femmine								
Nubili	87,0	10,7	2,3	100,0	70,5	15,0	14,5	100,0
Coniugate	16,7	70,0	13,3	100,0	8,1	44,7	47,3	100,0
Vedove	0,0	23,1	76,9	100,0	0,7	4,9	94,4	100,0
Divorziate	0,0	100,0	0,0	100,0	0,0	33,3	66,7	100,0

(a) I separati sono stati associati ai coniugati

5. Analisi in componenti principali ¹¹

Per sintetizzare l'intreccio delle relazioni di "interdipendenza" esistenti tra le diverse variabili in gioco, in un numero più ristretto di variabili si è effettuata un'analisi in componenti principali (ACP)

5.1 Cenni metodologici

L'ACP è una tecnica di analisi multivariata finalizzata alla rappresentazione sintetica di un numero elevato di variabili tra loro correlate, attraverso un numero più ridotto di nuove variabili tra loro ortogonali dette componenti principali (o assi), ognuna delle quali esprime una combinazione lineare delle variabili originarie ¹².

La prima componente estratta è quella a cui corrisponde la massima percentuale di varianza riprodotta; la seconda, ortogonale rispetto alla prima, esprimerà una quota di varianza riprodotta minore rispetto alla prima e così via fino ad estrarre tante componenti quante sono le variabili analizzate.

Naturalmente, basterà utilizzare solo le prime componenti estratte per avere un'efficace sintesi dei dati.

Per l'interpretazione e la valutazione dell'ACP ci si avvale di una serie di valori statistico-matematici.

L'autovalore di una componente estratta permette di valutare il suo potere di sintesi: più è grande l'autovalore maggiore è la percentuale di varianza riprodotta da quella componente.

La comunalità misura la percentuale di varianza di ciascuna variabile riprodotta dall'insieme delle componenti estratte (ovviamente la comunalità è uguale a 1 se si considerano tutte le possibili componenti, e minore di 1 se si lavora solo con le prime componenti estratte.)

Una volta stabilito il numero delle componenti da usare nell'analisi si passa a considerare il rapporto che c'è tra queste e le variabili.

A questo scopo, si utilizzano dei coefficienti chiamati *component loadings* (pesi delle componenti) che indicano appunto qual è la relazione lineare tra una variabile e una componente principale.

Si possono interpretare i *component loadings* come dei coefficienti di correlazione tra le variabili e le componenti estratte.

¹¹ Testo di Lidia Fagiolo

¹² Delvecchio F.,(1992), *Analisi statistica di dati multidimensionali*, Cacucci Editore, Bari.

Con la lettura di questi coefficienti si riesce ad interpretare semanticamente le componenti individuando dei sottogruppi di variabili che sono connessi in maniera più stretta a ciascuna di esse.

5.2 Principali risultati

I risultati ottenuti con l'ACP, effettuata con il programma SPAD (Système Portable Pour l'Analyse des Données), vengono illustrati di seguito.

La matrice dei dati su cui applicare l'ACP è stata costruita disponendo per riga le 19 regioni italiane (il Trentino non è stato incluso nell'analisi)¹³ e per colonna alcune delle variabili oggetto di studio, queste ultime sono elencate, con le relative etichette, nella tabella 13.

Tabella 13 – Le variabili utilizzate per ACP.

Etichette	Variabile
MASCHI	% decessi maschili
FEMMINE	% decessi femminili
MIN	% deceduti con meno di 30 anni
TRA	% deceduti tra 30 e 54 anni
MAGG	% deceduti con più di 55 anni
CEL	% deceduti celibi
CON	% deceduti coniugati
ALTR	% deceduti vedovi e divorziati
NOTTE	% incidenti notturni
GIORNO	% incidenti diurni
PROTSI	% uso protezioni obbligatorie
PROTNO	% non uso protezione obbligatorie
NATNNSPE	% incidenti di natura non specificata
SCCAV	% incidenti con scontro altro veicolo
SCP	% incidenti con scontro con pedone
PERCON	% incidenti per perdita controllo
ABIT	% incidenti nell'abitato
FABIT	% incidenti fuori abitato
INTER	% incidenti nelle intersezioni
NINTER	% incidenti nelle non intersezioni
ASCI	% incidenti su fondo stradale asciutto
NASCI	% incidenti su fondo stradale non asciutto

Estraendo le componenti principali si sono ottenuti i risultati presentati nella tabella 14; attraverso la sua lettura emerge che i primi tre fattori spiegano il 71% della varianza totale attraverso di essi si può, dunque, sintetizzare l'analisi dei dati con una perdita d'informazione contenuta.

¹³ Cfr. Cap. 2

Tabella 14 - Elenco delle componenti estratte.

Componente	Autovalori	Varianza spiegata (val. %)	Varianza spiegata (val % cumulata)
1	7,5843	34,47	34,47
2	4,3775	19,90	54,37
3	3,7768	17,17	71,54
4	1,8511	8,41	79,95
5	1,2657	5,75	85,71
6	0,8809	4,00	89,71
7	0,6404	2,91	92,62
8	0,4883	2,22	94,84
9	0,4065	1,85	96,69
10	0,3149	1,43	98,12
11	0,2795	1,27	99,39
12	0,0984	0,45	99,84
13	0,0351	0,16	100,00
14	0,0008	0,00	100,00
15	0,0000	0,00	100,00
16	0,0000	0,00	100,00
17	0,0000	0,00	100,00
18	0,0000	0,00	100,00
19	0,0000	0,00	100,00
20	0,0000	0,00	100,00
21	0,0000	0,00	100,00
22	0,0000	0,00	100,00

La tabella che segue riporta le variabili che hanno contribuito con più peso alla formazione delle tre componenti.

Tabella 15 – Pesì delle variabili (component loadings) sulle prime tre componenti

	Variabile	Etichetta	Peso
Comp. I	% incidenti diurni	GIOR	0,87
	% incidenti di natura non specificata	NATNNS	0,81
	% incidenti nelle non intersezioni	NINTER	0,80
	% decessi femminili	FEMMINE	0,71
	% deceduti coniugati	CON	0,67
	% deceduti oltre 55 anni	MAGG	0,61
	% incidenti su fondo stradale asciutto	ASCI	0,51
	% incidenti su fondo stradale non asciutto	NASCI	-0,51
	% decessi maschili	MASCHI	-0,71
	% incidenti con scontro altro veicolo	SCCAV	-0,76
	% incidenti per perdita controllo	PERCON	-0,78
	% incidenti nelle intersezioni	INTER	-0,80
	% incidenti notturni	NOTTE	-0,87
Comp. II	% non uso protezione obbligatorie	PROTNO	0,81
	% incidenti fuori dall'abitato	FABIT	0,74
	% deceduti con meno di 30 anni	MIN	0,56
	% decessi maschili	MASCHI	0,50
	% incidenti di natura non specificata	NATNN	0,40
	% deceduti oltre 55 anni	MAGG	-0,45
	% incidenti con scontro altro veicolo	SCCAV	-0,47
	% deceduti celibi	CEL	-0,49
	% decessi femminili	FEMMINE	-0,50
	% incidenti nell'abitato	ABIT	-0,74
% uso protezione obbligatorie	PROTSI	-0,81	
Comp. III	% incidenti su fondo stradale asciutto	ASCI	0,70
	% deceduti celibi	CEL	0,69
	% deceduti con meno di 30 anni	MIN	0,51
	% incidenti notturni	NOTTE	0,41
	% incidenti fuori dall'abitato	FABIT	0,40
	% incidenti nell'abitato	ABIT	-0,40
	% incidenti diurni	GIOR	-0,41
	% deceduti oltre 55 anni	MAGG	-0,43
	% deceduti coniugati	CON	-0,53
	% incidenti con scontro con pedone	SCP	-0,62
	% incidenti su fondo stradale non asciutto	NASCI	-0,70

Tra le diverse variabili che contribuiscono alla definizione della prima componente principale (quali ad esempio il sesso, la natura dell'incidente, le caratteristiche del fondo stradale...) assume particolare rilievo l'ora: si nota infatti una netta contrapposizione tra gli incidenti avvenuti nelle ore diurne e quelli avvenuti nelle ore notturne (tra le 22 e le 6 del mattino).

La seconda componente è caratterizzata dall'uso delle protezioni obbligatorie, dal sesso e dall'età delle vittime e dalla localizzazione degli incidenti che vede contrapposti quelli occorsi nell'abitato da quelli avvenuti, invece, fuori dall'abitato.

Risulta interessante l'analisi del terzo asse che sembra contrapporre in maniera abbastanza netta due tipologie di incidente stradale. Sul semiasse positivo troviamo le caratteristiche degli incidenti tipicamente giovanili: l'età delle vittime è inferiore ai trent'anni, gli incidenti avvengono di notte e fuori dall'abitato, e il fondo stradale asciutto fa presupporre che le cause dell'incidente siano da attribuire ad altri fattori (quali ad esempio l'eccessiva velocità).

Sul semiasse negativo troviamo, invece, le peculiarità degli incidenti che vedono coinvolti soprattutto i meno giovani: l'età delle vittime è oltre i cinquantacinque anni, gli incidenti avvengono durante le ore diurne e nel centro abitato, e il fondo stradale non asciutto porta a pensare che gli incidenti siano dovuti soprattutto a fattori esterni.

Questo asse sembra dunque contrapporre un diverso modo di utilizzare l'automobile: da una parte troviamo l'uso legato principalmente allo svago e dall'altra l'uso legato alle necessità quotidiane.

Non è facile riuscire a "denominare" le componenti in base alle variabili che le caratterizzano ma questo non impedisce di raggiungere un obiettivo forse più importante in questo ambito: evidenziare una "caratterizzazione" a livello regionale degli incidenti stradali occorsi nell'anno 2000.

Il programma SPAD permette, infatti, le proiezioni dei singoli casi (le 19 regioni italiane) stabilendo per ognuno le coordinate sulle componenti principali, i suoi contributi ad ogni componente principale e i coseni quadrati che si interpretano come qualità della rappresentazione che ogni componente fornisce per ogni caso.

Attraverso tali dati è dunque possibile stabilire quanto il singolo caso è in "relazione" con le diverse componenti e conseguentemente con le variabili ad esse correlate.

La tabella che segue riporta questi valori:

Tabella 16 - Matrice delle coordinate, dei contributi e dei coseni quadrati dei casi sulle tre componenti estratte.

	Coordinate			Contributi			Coseni quadrati		
	Comp.I	Comp.II	Comp.III	Comp.I	Comp.II	Comp.III	Comp.I	Comp.II	Comp.III
Abruzzo	-0,15	1,09	0,71	0,0	1,4	0,7	0,00	0,09	0,04
Basilicata	2,55	1,47	3,25	4,5	2,6	14,7	0,30	0,10	0,48
Calabria	2,87	1,25	-2,30	5,7	1,9	7,4	0,41	0,08	0,27
Campania	0,72	1,56	-2,45	0,4	2,9	8,4	0,04	0,18	0,46
Emilia	-3,1	-1,59	-1,07	6,7	3,1	1,6	0,64	0,17	0,08
Friuli	-4,8	-1,94	0,05	16,0	4,5	0,0	0,63	0,10	0,00
Lazio	-1,03	1,92	0,36	0,7	4,4	0,2	0,14	0,50	0,02
Liguria	3,13	0,43	-4,29	6,8	0,2	25,7	0,28	0,01	0,54
Lombardia	-4,34	-1,45	-0,53	13,1	2,5	0,4	0,66	0,07	0,01
Marche	-1,01	-0,23	2,54	0,7	0,1	9,0	0,06	0,00	0,39
Molise	3,27	2,64	-0,06	7,4	8,4	0,0	0,42	0,27	0,00
Piemonte	-3,61	-0,13	0,56	9,1	0,0	0,4	0,67	0,00	0,02
Puglia	-0,3	2,66	2,21	0,1	8,5	6,8	0,01	0,40	0,28
Sardegna	0,93	-0,06	2,44	0,6	0,0	8,3	0,05	0,00	0,34
Sicilia	1,73	0,30	1,95	2,1	0,1	5,3	0,21	0,01	0,27
Toscana	-1,62	0,09	-2,24	1,8	0,0	7,0	0,16	0,00	0,31
Umbria	1,56	-0,56	-1,20	1,7	0,4	2,0	0,17	0,02	0,10
Val d'Aosta	5,31	-6,99	0,92	19,6	58,7	1,2	0,36	0,62	0,01
Veneto	-2,1	-0,44	-0,85	3,1	0,2	1,0	0,68	0,03	0,11

Vediamo dunque la proiezione sui due piani fattoriali (ottenuti dalla combinazione delle tre componenti) delle variabili e delle regioni.

Grafico 3 – Proiezione delle variabili sul primo piano fattoriale (comp. I e II)

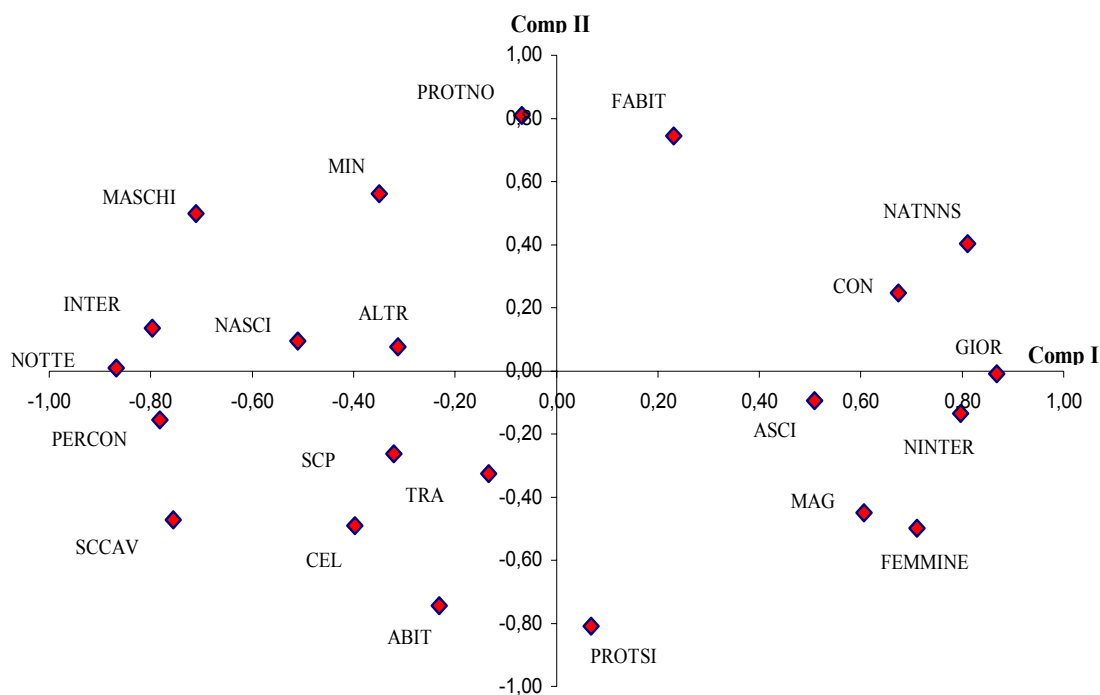
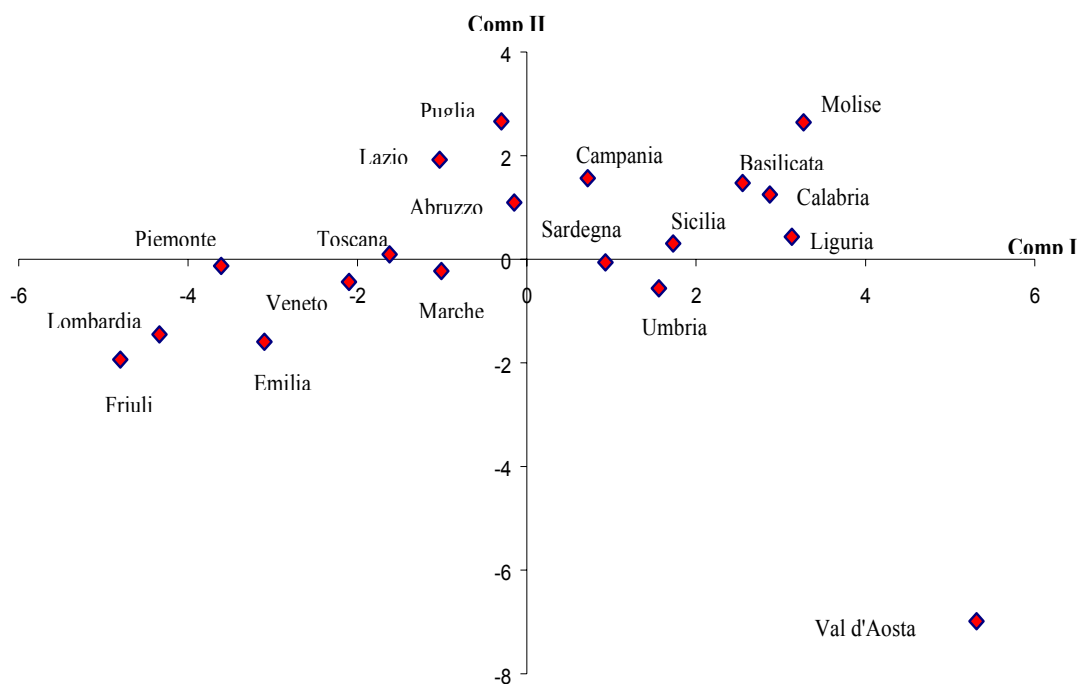


Grafico 4 – Proiezione delle unità sul primo piano fattoriale (comp. I e II)



La Calabria, la Liguria, la Lombardia, il Friuli, l'Emilia e il Piemonte sono le regioni che danno un contributo elevato alla formazione della prima componente (la grandezza dei contributi è determinata dal rapporto tra 100, che rappresenta la somma dei contributi di tutti i casi alle componenti, e 19, che è il numero delle regioni. In questo caso il contributo medio di ogni regione alle varie componenti è pari a 5).

Le prime due regioni si posizionano sul semiasse positivo; questo sta a significare che sono principalmente caratterizzate da incidenti accaduti di giorno, che vedono coinvolte donne coniugate di età superiore ai 54 anni.

Le rimanenti regioni sono sul semiasse negativo e sono dunque contraddistinte da sinistri avvenuti nelle ore notturne, che vedono coinvolti uomini alle prese con fondo stradale non asciutto e presenza di intersezioni.

Contribuiscono alla formazione della seconda componente la Puglia, il Lazio e la Val d'Aosta: le prime due si posizionano sul semiasse positivo (sono interessante dunque da incidenti che accadono fuori dal centro abitato e che vedono coinvolti maschi sotto i trent'anni che non fanno uso delle protezioni obbligatorie) e l'altra sul versante opposto (incidenti che accadono nel centro abitato e che interessano donne sopra i 55 anni che indossano i dispositivi di sicurezza).

Il Molise con comportamento differenziato rispetto alle altre regioni fornisce contributi pressoché simili alla formazione dei due assi e si colloca nel primo quadrante, viene, quindi, caratterizzato dalle variabili che definiscono i due semiassi positivi.

Per quanto riguarda invece il secondo piano fattoriale (composto dalla prima e terza componente) le regioni che contribuiscono alla formazione della terza componente sono: la Basilicata, la Sardegna, la Sicilia, le Marche, la Calabria, la Campania, la Liguria e la Toscana. Le prime quattro posizionandosi sul semiasse positivo sono interessate da incidenti tipicamente giovanili, le altre, invece, da quelli che riguardano soprattutto persone adulte.

Grafico 5 – Proiezione delle variabili sul secondo piano fattoriale (comp. I e III)

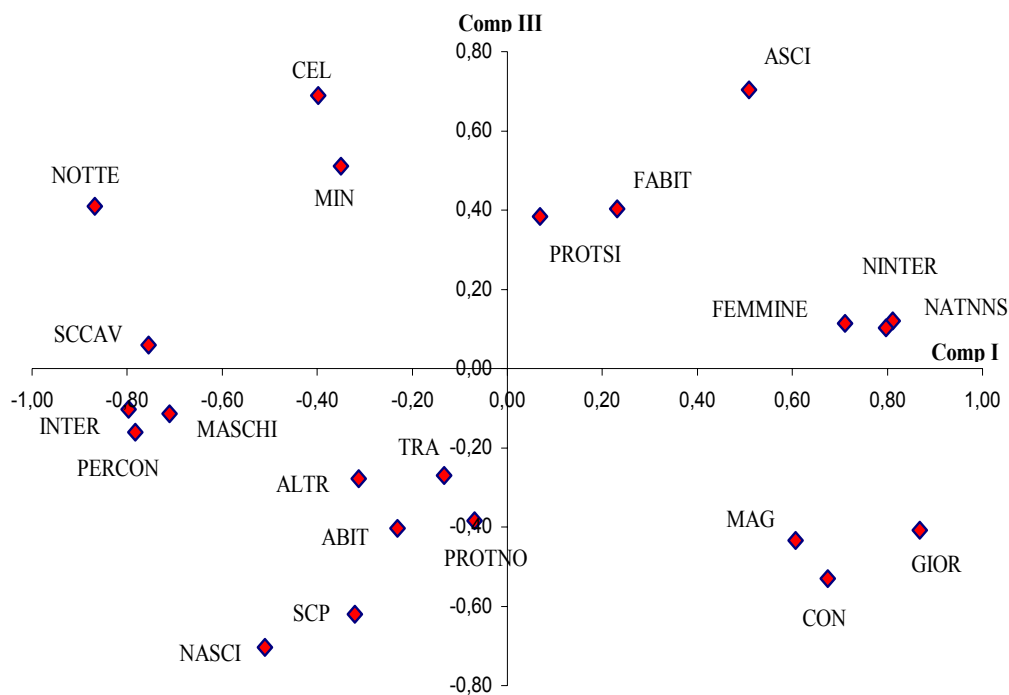
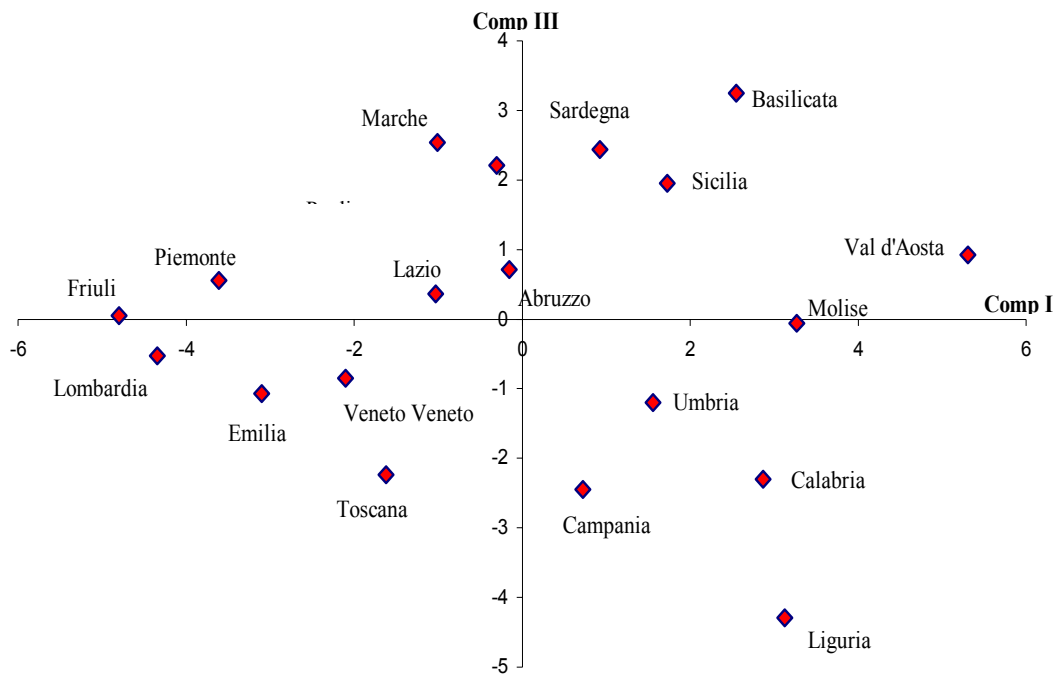


Grafico 6 – Proiezione delle unità sul secondo piano fattoriale (comp. I e III)



6. Conclusioni¹⁴

L'Italia, come gli altri paesi dell'Unione Europea, si è posta l'obiettivo di ridurre del 40 per cento entro il 2010 il numero di morti e di feriti causati da incidenti stradali.

Questo rende sempre più necessario disporre di informazioni attendibili che permettono di monitorare l'effettivo livello di sicurezza stradale.

Per effetto degli incidenti stradali, la nostra società è soggetta ogni anno ad un costo che non ha confronti con qualsiasi altro fenomeno.

Tale problematica va, quindi, analizzata nella sua totalità cogliendo le molteplici peculiarità che la contraddistinguono.

Per un miglioramento della sicurezza stradale è necessario, perciò, operare su più fronti con un'azione coordinata negli interventi in quanto le competenze necessarie per affrontare questa materia sono innumerevoli e difficilmente sono concentrate in un'unica Istituzione.

L'attuale base informativa sull'incidentalità stradale è ampia ma al tempo stesso molto frammentaria; le informazioni statistiche sul settore costituiscono un insieme estremamente eterogeneo poiché provengono da molteplici fonti, alcune di natura amministrativa ed altre a carattere epidemiologico.

Questo comporta che i dati disponibili presentino talvolta livelli di qualità diversi e non del tutto comparabili tra loro, in quanto si riferiscono a rilevazioni che in alcuni casi adottano definizioni, classificazioni, nomenclature e metodologie diverse, risentono di un'impostazione indirizzata ad analizzare la sicurezza stradale con un approccio mirato; mentre i bisogni informativi portano alla necessità di adottare una visione integrata dei fenomeni.

La rilevazione Istat-ACI costituisce la fonte di informazioni più ampia sul fenomeno ma accentra unicamente l'attenzione sull'evento incidente mentre sembra necessario affiancare a queste indicazioni quelle relative all'ambiente stradale, agli esiti dell'incidente, alle lesioni riportate dagli infortunati, ai costi ospedalieri, ecc. provenienti anche da altre indagini.

La strada su cui si muove l'Istat è quella del miglioramento dell'informazione statistica nel settore dell'incidentalità stradale e per far questo si pone l'obiettivo di collegare in modo sistematico le statistiche sugli incidenti stradali alle altre statistiche, una fra tutte quella sui dimessi ospedalieri e quella sulle cause di morte.

Le informazioni fornite da queste statistiche contribuiscono se lette in una chiave unitaria, ossia, in forma integrata, ad identificare quali sono i principali fattori di rischio e, successivamente, forniscono i dati di partenza per definire una corretta politica di prevenzione.

¹⁴ Testo di Raffaella Amato e Silvia Bruzzone

Il lavoro svolto rappresenta un primo tentativo di collegamento tra le statistiche sugli incidenti stradali e quelle sulle cause di morte, sulla base di operazioni di record linkage tra i due diversi archivi.

Nonostante ci siano, evidentemente, dei limiti che non consentono di generalizzare i risultati ed estenderli all'intera popolazione, il presente lavoro costituisce sicuramente un punto di partenza per sviluppi futuri.

Bibliografia

COMMISSIONE EUROPEA. Promuovere la sicurezza stradale nell'Unione Europea: il programma 1997-2001, Bruxelles, 09/04/1997, COM (1997), 131 def.

Giustiniani M., Bagarini M., Cresca A., Granata O., Dosi G., Taggi F., *L'uso dei dispositivi di sicurezza in Italia: i dati dell'Osservatorio Nazionale (sistema Ulisse)*. Atti della Consulta Nazionale della Sicurezza Stradale, III sessione plenaria, CNEL, 3 Luglio 2002.

Istat, Annuari n.16 (2004), *Cause di morte. Anno 2000*, Roma.

Istat, Informazioni n.17 (1999), *Statistiche sugli incidenti stradali. Anno 2000*, Roma.

Istat, Argomenti n.7 (1997), *Gli incidenti stradali negli anni '90. Anno 1997*, Roma.

Istat-Iss (s.d.) *La mortalità in Italia nel periodo 1970-1992: evoluzione e geografia*, ISTAT-ISS, Istituto Poligrafico e Zecca dello Stato, Roma.

Istat, Metodi e Norme (1998, 6^a ristampa), *Classificazione delle malattie, traumatismi e cause di morte, 9^a revisione 1975*. Roma.

Istat, Metodi e Norme n.16 (2003), *Metodi statistici per il record linkage*. Roma.

Ministero delle Infrastrutture e dei Trasporti, *Indirizzi generali e linee-guida di attuazione del Piano Nazionale della Sicurezza Stradale*, Marzo 2000.

Ministero delle Infrastrutture e dei Trasporti, *Piano Nazionale della Sicurezza Stradale: Azioni Prioritarie*, Marzo 2002.

Putignano C., (cura di), Metodi e Norme, Serie B n.28 (1990), *Istruzioni per la rilevazione degli incidenti stradali*. ISTAT, Roma.

Putignano C., *L'incidentalità stradale in Italia: dimensione del fenomeno e confronti internazionali*, Rivista trimestrale del Centro studi sui sistemi di trasporto, n.2, Aprile-Giugno 1995.

Taggi F. (a cura di), (2003) *Aspetti sanitari della sicurezza stradale (secondo rapporto sul progetto DATIS)*. ISS, Roma.

Taggi F., di Cristofaro Longo G., (2001) *I dati socio-sanitari della sicurezza stradale (Progetto DATIS)*, ISS- Ministero delle Infrastrutture e dei Trasporti, Roma.

APPENDICE A

Schema della metodologia applicata ¹⁵

In linee generali, la metodologia presentata consiste nella reiterazione di un programma di merge, applicato in diversi cicli su file opportunamente modificati.

In seguito ad un monitoraggio congiunto delle due popolazioni si è notato che molte unità presentano le stesse modalità delle variabili comuni precedentemente illustrate, ma un'età che differisce nell'intervallo compreso tra -2 e +2 anni.

Poiché le informazioni anagrafiche contenute nell'indagine sulle cause di morte sono fornite dall'Ufficiale di Stato Civile si ritiene che siano più attendibili rispetto a quelle dell'indagine sugli incidenti stradali dove sono fornite dall'Autorità di Polizia.

Alla luce di tale considerazione, per stabilire una compatibilità con i dati relativi alle cause di morte, si è modificata la variabile età (maggiorandola e minorandola rispettivamente di una e di due unità) nel file degli incidenti stradali.

Tale trasformazione si è effettuata sulle unità che di volta in volta non trovano corrispondenza con quelle presenti nelle cause di morte.

La variabile di linkage non è univoca per ogni record dunque si è utilizzato un artificio per separare in file diversi record originali dai relativi dopponi (ossia quelli che presentano informazioni relative alle variabili comuni uguali tra loro).

A tal fine si sono costruite due variabili denominate primo ed ultimo che identificano rispettivamente la posizione iniziale e finale nella chiave di linkage.

I valori assunti da queste variabili sono pari ad 1 o 0, a seconda che il record è originale o possiede delle ripetizioni:

a. se primo = 1 e ultimo =1 —————> record originale senza possibilità di dopponi

b. se primo = 1 e ultimo =0 —————> record originale con successivi dopponi



a.+ b. = **File dei Record Originali**

¹⁵ Testo di Valentina Delmonte e Lidia Fagiolo.

c. se primo = 0 e ultimo =0 —————> record doppi di b. che potrebbero ripetersi
al loro interno



c. = **File Doppi1**

d. se primo = 0 e ultimo =1 —————> record doppi di c., quindi di b., ma che non si
ripetono al loro interno



d. = **File Doppi2**

I risultati ottenuti applicando la procedura illustrata sono i seguenti:

Cause di morte

- a. Record Originali = 6099
- c. Doppi1 = 0
- e. Doppi2 = 27

Incidenti stradali

- b. Record Originali = 5592
- d. Doppi1 = 37
- f. Doppi2 = 373

Per prima cosa si è condotta un'analisi di linkage dei file contenenti i record originali delle due indagini in due cicli:

- nel primo ciclo si è utilizzata la chiave di linkage precedentemente descritta;
- nel secondo ciclo si è utilizzata una nuova chiave non contenente il comune del decesso, in modo da ridurre i vincoli per l'associazione dei record residui.

La procedura descritta è illustrata dal seguente schema:

1° CICLO

Chiave di linkage = meseinc || giorninc || comdec || provdec || sesso || eta

Incidenti stradali \longrightarrow **Cause di morte**
a. File record originali (6099) b. File record originali (5592)



1) Merge tra file a. e file b. \longrightarrow **1610 record linkati**

2) Eliminazione da a. e b. dei 1610 record ottenuti al passo 1)

g. incidenti stradali	4489	} Ammontare restante delle due popolazioni
h. cause di morte	3982	

3) Modifica della variabile età nel file g. (eta=eta+1)

4) Merge tra file g. modificato e file h. \longrightarrow **46 record linkati**

5) Eliminazione da g. e h. dei 46 record ottenuti al passo 4)

i. incidenti stradali	3936	} Ammontare restante delle due popolazioni
l. cause di morte	4443	

6) Modifica della variabile età nel file i. (eta=eta-1)

7) Merge tra file i modificato e file l. \longrightarrow **858 record linkati**

8) Eliminazione da i. e l. degli 858 record ottenuti al passo 7)

m. incidenti stradali	3078	}	Ammontare restante delle due popolazioni
n. cause di morte	3585		

9) Modifica della variabile età nel file m. ($\text{eta}=\text{eta}-2$)

10) Merge tra file m modificato e file n. \longrightarrow **16 record linkati**

11) Eliminazione da m. e n. dei 16 record ottenuti al passo 10)

o. incidenti stradali	3062	}	Ammontare restante delle due popolazioni
p. cause di morte	3569		

12) Modifica della variabile età nel file o. ($\text{eta}=\text{eta}+2$)

13) Merge tra file o. modificato e file p. \longrightarrow **3 record linkati**

14) Eliminazione da o. e p. dei 3 record ottenuti al passo 13)

q. incidenti stradali	3059	}	\longrightarrow Record residui
r. cause di morte	3566		



Totale record linkati = 1610 + 46 + 858 + 16 + 3 = 2533

2° CICLO

In questo secondo ciclo, dopo aver costruito la nuova chiave di linkage, che non tiene conto del comune di incidente:

Chiave di linkage = meseinc || giorninc || provdec || sesso || eta

si è ripetuta la metodologia precedentemente descritta considerando i record residui ottenuti dal primo ciclo:

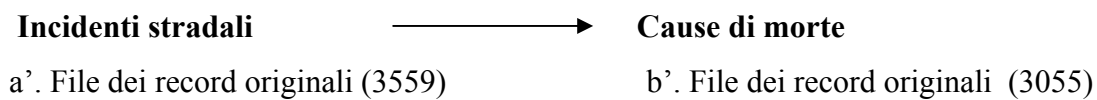
Cause di morte

- a.' Record Originali = 3559
- c.' Doppi1 = 0
- e.' Doppi2 = 7

Incidenti stradali

- b.' Record Originali = 3055
- d.' Doppi1 = 0
- f.' Doppi2 = 4

A questo punto si è applicata la procedura precedentemente descritta:



1) Merge tra file a'. e file b'. \longrightarrow **664 record linkati**

2) Eliminazione da a'. e b'. dei 664 record 1)

g'. incidenti stradali	2391	} Ammontare restante delle due popolazioni
h'. cause di morte	2895	

3) Modifica della variabile età nel file g'. (eta=eta+1)

4) Merge tra file g' modificato e file h'. \longrightarrow **13 record linkati**

5) Eliminazione da g'. e h'. dei 13 record ottenuti al passo 4)

i'. incidenti stradali	2378	}	Ammontare restante delle due popolazioni
l'. cause di morte	2882		

6) Modifica della variabile età nel file i'. (eta=eta-1)

7) Merge tra file i' modificato e file l'. \longrightarrow **347 record linkati**

8) Eliminazione da i'. e l'. dei 347 record ottenuti al passo 7)

m'. incidenti stradali	2031	}	Ammontare restante delle due popolazioni
n'. cause di morte	2535		

9) Modifica della variabile età nel file m'. (eta=eta+2)

10) Merge tra file m' modificato e file n'. \longrightarrow **1 record linkato**

11) Eliminazione da m'. e n'. del record ottenuto al passo 10)

o'. incidenti stradali	2534	}	Ammontare restante delle due popolazioni
p'. cause di morte	2030		

12) Modifica della variabile età nel file o'. (eta=eta-2)

13) Merge tra file o' modificato e file p'. \longrightarrow **12 record linkati**

14) Eliminazione da o'. e p'. dei 12 record ottenuti al passo 13)

q'. incidenti stradali	2018	\longrightarrow	Record residui
r'. cause di morte	2522		



Totale record linkati = $664 + 13 + 347 + 1 + 12 = 1037$

Successivamente si è passati ad analizzare i file contenenti i record doppi con due nuovi cicli in cui si sono tenute in considerazione le due chiavi di link descritte:

3° CICLO

Chiave di linkage= meseinc || giornincc || comdec || provdec || sesso || eta

Incidenti stradali	\longrightarrow	Cause di morte
c. File dei doppi1 (0)		d. File dei doppi1 (37)
e. File dei doppi2 (27)		f. File dei doppi2 (373)



! Non è possibile Merge tra file c. e d. (Nel file c. non ci sono record)

1) Merge tra file e. e file f. \longrightarrow **7 record linkati**

2) Eliminazione da e. e f. dei 7 record ottenuti al passo 1)

→	g. incidenti stradali	366	}	Ammontare restante delle due popolazioni
	h. cause di morte	20		

3) Modifica della variabile età nel file g. ($eta=eta-1$)

4) Merge tra file g. modificato e file h. → **6 record linkati**

5) Eliminazione da g. e h. dei 6 record ottenuti al passo 4)

→	i. incidenti stradali	360	}	Ammontare restante delle due popolazioni
	l. cause di morte	14		

6) Modifica della variabile età nel file i. ($eta=eta+1$)

7) Merge tra file modificato i. e file l. → **0 record linkati**

! Con $eta=eta-2$ $eta=eta+2$ → Nessun risultato



Totale record Linkati = 7 + 6 = 13

4° CICLO

Chiave di linkage= meseinc || giorninc || provdec || sesso || eta

1) Modifica età nel file i. (eta=eta-1)

2) Merge tra file modificato i e file l \longrightarrow **2 record linkati**

3) Eliminazione da i. e l. dei 2 record ottenuti al passo 2)

\longrightarrow	m. incidenti stradali	358	\longrightarrow	Record Residui
	n. cause di morte	12		

! Con eta=eta+1 eta=eta-2 eta=eta+2 \longrightarrow Nessun risultato

! Il tentativo di merge tra i file contenenti i record doppiati ottenuti nel secondo ciclo non ha fornito alcun risultato.



Numerosità totale del sottoinsieme = 2533+1037+13+2 = 3585

APPENDICE B

MODELLI DI RILEVAZIONE