



Università degli Studi di Salerno
Dipartimento di Informatica
Corso di Laurea Triennale in Informatica

Progetto Calcolo Probabilità Statistica Matematica
(CPSM)

Indagine Statistica sulle Morti in incidenti stradali

Tozza Gennaro Carmine
Matricola: 0512120382

Anno Accademico 2024-2025

Indice

1	Introduzione	2
1.1	Problematica	2
1.2	Scopo del progetto	2
2	Tabelle delle frequenze	4
3	Rappresentazione dei dati mediante grafici	6
4	Indici di posizione	8
4.1	Media campionaria	8
4.2	Mediana campionaria	8
4.3	Moda campionaria	8
5	Indici di variabilità	9
5.1	Varianza campionaria	9
5.2	Deviazione standard campionaria	9
5.3	Scarto medio assoluto	9
5.4	Ampiezza del campo di variazione	9
5.5	Coefficiente di variazione	9
6	Indici di forma	10
6.1	Indice di asimmetria	10
6.2	Indice di curtosi	10

Capitolo 1

Introduzione

1.1 Problematica

Gli **incidenti stradali** costituiscono una delle principali emergenze di sanità pubblica, in quanto responsabili ogni anno di un elevato numero di decessi, in particolare tra i giovani, e di gravi conseguenze in termini di disabilità temporanee e permanenti, oltre al drammatico impatto umano e psicologico sulle vittime e sulle loro famiglie.

1.2 Scopo del progetto

Il progetto consiste in un'indagine statistica ¹sugli incidenti stradali verificatisi sulla rete stradale del territorio nazionale, tra il 2010 e il 2023 verbalizzati da un'autorità di Polizia o dai Carabinieri, avvenuti su una strada aperta alla circolazione pubblica e che hanno causato lesioni a persone, morti (entro il 30° giorno) e/o feriti, con il coinvolgimento di almeno un veicolo.

La rilevazione è condotta correntemente dall'Istat, con la compartecipazione dell'ACI e di numerosi Enti pubblici istituzionali, è a carattere totale e a cadenza mensile (inserita tra le rilevazioni di interesse pubblico nel Programma Statistico Nazionale - PSN - IST00142).

Per l'analisi dei dati è stato scelto l'ambiente di calcolo statistico **R**.

R fornisce un'ampia varietà di tecniche statistiche (modellazione lineare e non lineare, test statistici classici, analisi delle serie temporali, classificazione, ...) e grafiche ed è altamente estensibile.

Uno dei punti di forza di R è la facilità con cui possono essere prodotti grafici ben progettati e di qualità per la pubblicazione, compresi simboli matematici e

¹<https://siqua.istat.it/SIQual/visualizza.do?id=7777778&refresh=true&language=IT>

formule se necessario.

Per semplificare l'analisi, si è scelto di lavorare non sull'intero dataset, ma su un sottoinsieme filtrato di dati, relativo alle morti per incidenti stradali che riguardano solo i conducenti di età compresa tra i 21 e i 24 anni.

Per approfondire l'analisi con dati dettagliati e specifici, è possibile consultare il dataset direttamente sul sito dell'ISTAT al seguente link: https://esploradati.istat.it/databrowser/#/it/dw/categories/IT1,Z0810HEA,1.0/HEA_ROAD/IT1,41_270_DF_DCIS_MORTIFERITISTR1_1,1.0

Il dataset in formato CSV è stato ottenuto dalla fonte ISTAT tramite il link indicato, assicurando così l'affidabilità dei dati.

Il formato scelto (CSV) permette un'agevole manipolazione dei dati, essendo compatibile con la maggior parte dei software statistici e dei fogli di calcolo, ottimizzando l'analisi e la visualizzazione delle informazioni.

Intersezione	'10	'11	'12	'13	'14	'15	'16	'17	'18	'19	'20	'21	'22	'23
Incrocio	55	45	42	23	34	24	28	25	23	13	15	16	21	10
Rotatoria	7	4	2	5	5	1	5	3	1	2	1	2	–	3
Rettilineo	92	102	89	82	96	95	66	70	60	72	66	69	69	74
Curva	58	51	48	46	45	44	41	36	38	42	23	40	27	38
Dosso/Pend.	2	7	2	4	2	3	5	1	2	3	2	4	3	1
Galleria	1	1	1	1	–	1	1	1	–	–	3	1	–	–
Tot.	215	210	184	161	182	168	146	136	124	132	110	132	120	126

Capitolo 2

Tabelle delle frequenze

```
1 # inclusione librerie
2 library("tidyverse")
3 require("tidyverse")
4 library("dplyr")
5
6 # viene caricato il dataset
7 dati <- read.csv("dati_istat.csv")
8
9 # filtraggio dati
10 dati <- dati %>% select(Intersezione, TIME_PERIOD,
11   Osservazione)
12 dati <- dati %>%
13   filter(Intersezione != "Totale")
14
15 # 1. Frequenze assolute per intersezione dell'incidente
16 freq_assolute <- aggregate(Osservazione ~ Intersezione, data
17   =dati, sum)
18
19 colnames(freq_assolute) <- c("Intersezione", "Frequenza_
20   Assoluta")
21 print(freq_assolute)
22
23 # 2. Frequenze relative per intersezione dell'incidente
24 totale <- sum(freq_assolute$Frequenza_Assoluta)
25 freq_assolute$Frequenza_Relativa <- freq_assolute$Frequenza_
26   Assoluta / totale
27 print(freq_assolute)
28
29 # 3. Frequenze cumulate assolute
30 freq_assolute <- freq_assolute[order(-freq_assolute$
31   Frequenza_Assoluta),]
32 freq_assolute$Frequenza_Cumulata_Assoluta <- cumsum(freq_
33   assolute$Frequenza_Assoluta)
34 print(freq_assolute)
```

```

28
29 # 4. Frequenze cumulate relative
30 freq_assolute$Frequenza_Cumulata_Relativa <- freq_assolute$
    Frequenza_Cumulata_Assoluta / totale
31 print(freq_assolute)

```

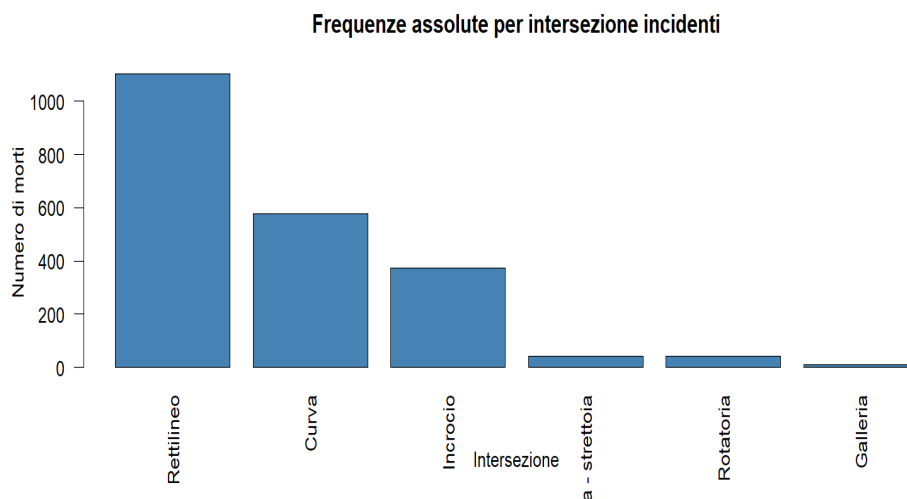
Output:

Intersezione	Freq. Assoluta	Freq. Relativa	Freq. Cum. Assoluta	Freq. Cum. Relativa
Rettilineo	1102	0.5135	1102	0.5135
Curva	577	0.2689	1679	0.7824
Incrocio	374	0.1743	2053	0.9567
Dosso - pendenza - strettoia	41	0.0191	2094	0.9758
Rotatoria	41	0.0191	2135	0.9949
Galleria	11	0.0051	2146	1.0000

Capitolo 3

Rappresentazione dei dati mediante grafici

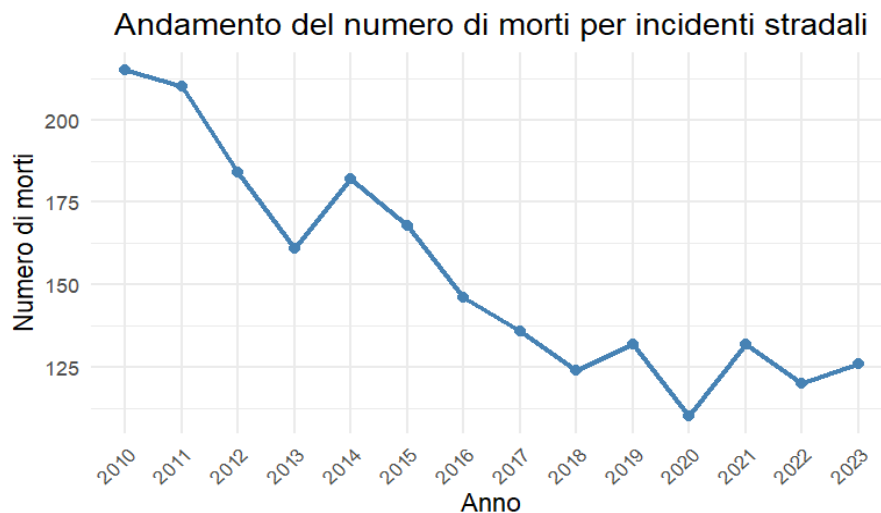
```
1 # Visualizzazione grafica
2 barplot(freq_assolute$Frequenza_Assoluta,
3         names.arg = freq_assolute$Intersezione,
4         main = "Frequenze assolute per intersezione incidenti",
5         xlab = "Intersezione",
6         ylab = "Numero di morti",
7         col = "steelblue",
8         las = 2) # Etichette verticali
```



```

1 # Converti TIME_PERIOD in fattore per mantenere l'ordine
  originale
2 dati$TIME_PERIOD <- factor(dati$TIME_PERIOD, levels = unique
  (dati$TIME_PERIOD))
3
4 # Calcola il numero totale di morti per anno
5 morti_per_anno <- dati %>%
6   group_by(TIME_PERIOD) %>%
7   summarise(Totale_Morti = sum(Osservazione, na.rm = TRUE))
8
9 # Crea il grafico a linee con tutti gli anni visibili
10 ggplot(morti_per_anno, aes(x = TIME_PERIOD, y = Totale_Morti
  , group = 1)) +
11   geom_line(color = "steelblue", size = 1) +
12   geom_point(color = "steelblue", size = 2) +
13   labs(title = "Andamento del numero di morti per incidenti
  stradali",
14         x = "Anno",
15         y = "Numero di morti") +
16   theme_minimal() +
17   theme(plot.title = element_text(hjust = 0.5),
18         axis.text.x = element_text(angle = 45, hjust = 1,
19                                     size = 8)) + # Riduci dimensione testo
19   scale_x_discrete(breaks = levels(morti_per_anno$TIME_
20                           PERIOD)) # Mostra tutti i valori

```



Capitolo 4

Indici di posizione

4.1 Media campionaria

```
1 media_generale <- mean(dati$Osservazione, na.rm = TRUE)
2 print(paste("Media_campionaria_generale:", round(media_
             generale, 2)))
```

Output: "Media campionaria generale: 27.51"

4.2 Mediana campionaria

```
1 mediana_generale <- median(dati$Osservazione, na.rm = TRUE)
2 print(paste("Mediana_generale:", mediana_generale))
```

Output: "Mediana generale: 15.5"

4.3 Moda campionaria

Capitolo 5

Indici di variabilità

- 5.1 Varianza campionaria
- 5.2 Deviazione standard campionaria
- 5.3 Scarto medio assoluto
- 5.4 Ampiezza del campo di variazione
- 5.5 Coefficiente di variazione

Capitolo 6

Indici di forma

6.1 Indice di asimmetria

6.2 Indice di curtosi