

Quanti feriti a Roma!!

Big data
Metodi statistici per la società della conoscenza

a.a. 2020/2021

Un breve rapporto sul numero di feriti riportati negli incidenti stradali rilevati a Roma.
Premessa. Il seguente percorso guidato di esplorazione aiuta a rendere operative alcune delle abilità sulla manipolazione di strutture di dati come il `data.frame` o il `data.table` esposte durante il corso. I dati sono stati prelevati dal portale Open Data di Roma Capitale <https://dati.comune.roma.it> e si basano su registrazioni di carattere amministrativo. I dati sono stati quindi aggregati in un unico file ASCII denominato "Incidenti-a-Roma.csv"

1. Importare i dati presenti nel file "Incidenti-a-Roma.csv" in un oggetto denominato `incidenti`
2. Quale variabile contiene il maggior numero di dati mancanti?
3. Inserire nel dataset una nuova variabile che contenga l'informazione sul solo giorno di rilevazione rimuovendo l'informazione sull'orario.
4. Sfruttare la nuova colonna per derivare un nuovo dataset che sarà denominato `dati_feriti_giornalieri` in cui sia presente per ciascun giorno di rilevazione il numero totale di feriti registrati in tutti gli incidenti rilevati in quel giorno.
5. Per quanti giorni distinti hai a disposizione il numero di feriti totali giornalieri rilevati negli incidenti stradali a Roma?
6. Per quanti anni sono stati calcolati i feriti totali giornalieri negli incidenti?
7. Gli anni sono consecutivi?
8. Creare una nuova variabile del nuovo dataset `dati_feriti_giornalieri` che contenga l'informazione sul solo giorno della settimana
9. Deriva una tabella che contenga il numero totale complessivo in quegli anni di feriti per ciascun giorno della settimana ad iniziare dal lunedì fino alla domenica.
10. Sulla base del nuovo dataset `dati_feriti_giornalieri` quali delle seguenti funzioni può essere utile per ottenere una tabella o `data.frame` con il numero di feriti complessivamente registrato in ciascuno dei 7 giorni della settimana iniziando dal giorno Lunedì fino alla Domenica in modo automatico [senza richiedere la riorganizzazione manuale dell'ordine delle righe]? (mostrare come)
 - `sort(...)`
 - `which.max(...)`
 - `rank(...)`

- `factor(...)`
- `unique(...)`
- `aggregate(...)`

11. Qual è il giorno della settimana con maggior numero di feriti totale?

12. Qual è il giorno della settimana con il minor numero di feriti per incidente?

13. Considera il numero di feriti in ciascun incidente rilevato a Roma nel periodo da 1 ottobre 2010 al 15 dicembre 2010 nella fascia oraria 14:00 14:59 della domenica. Il dataset è contenuto nel file denominato "feriti_domenica_14.csv". Utilizza il test chi-quadrato `chisq.test(...)` per verificare l'ipotesi che la distribuzione del numero di feriti per incidente in quella fascia oraria sia compatibile con un modello di distribuzione teorica fornita dalla seguente distribuzione di probabilità: (Tale distribuzione teorica è stata derivata dal modello di Poisson con parametro media uguale alla media calcolata sulla base delle osservazioni.)

```
p_teorica <- c(0.568, 0.322, 0.091, 0.017, 0.002)
```

Rispondi alle seguenti domande:

- quanto vale il p-value del test?
- a quali conclusioni ci porta?

14. Consegna la tua elaborazione come file `Cognome-Nome.R` o, meglio ancora, come file `Cognome-Nome.Rmd` (RMarkdown).