

Tema d'esame di Statistica e analisi dei dati

Prova scritta del 26 gennaio 2016

Esercizio 0

1. Si tracci un grafico indicativo (senza utilizzare comandi R) della funzione di ripartizione di una variabile casuale uniforme discreta che assume valori nell'insieme $\{2, 3, 4, \dots, 14\}$.

Sia X una variabile casuale geometrica di parametro p che assume valori nell'insieme $\{0, 1, 2, 3, \dots\}$.

2. Si tracci un grafico indicativo (senza utilizzare comandi R) della funzione massa di probabilità di X .
3. Si esprima, in funzione di p , il valore atteso $E(X)$.
4. Si esprima p in funzione di $E(X)$.
5. Si esprima, in funzione di $E(X)$, la varianza di X .

Collegatevi al sito upload.di.unimi.it e selezionate l'esame di *Statistica e analisi dei dati*. Scaricate il file `Comune.Bergamo_-_Incidenti_stradali.csv`, che contiene oltre 26000 descrizioni di incidenti avvenuti tra il 2002 e il 2014 nel comune di Bergamo (Fonte: Regione Lombardia). Il file contiene una riga per incidente, i cui campi, separati da virgole, sono i seguenti:

- Protocollo, numero identificativo del verbale;
- Anno, anno dell'incidente;
- Data, data dell'incidente;
- Ora, ora dell'incidente;
- Minuto, minuto dell'incidente;
- Localita, luogo dell'incidente;
- NaturaIncidente, tipo di incidente;
- N_Illesi, numero di persone illese;
- N_Feriti, numero di feriti;
- N_Riservata, campo riservato;
- N_Morti, numero di morti;

- Pedoni, coinvolgimento di pedoni;
- Velocipedi, coinvolgimento di velocipedi;
- Ciclomotori_Motocicli, coinvolgimento di ciclomotori o motocicli;
- Mezzi_Pesanti, coinvolgimento di mezzi pesanti;
- Localizzazione, latitudine e longitudine del luogo dell'incidente.

Esercizio I

1. Qual è il numero n di record nel dataset?
2. Indicare quali campi contengono valori scalari, quali valori ordinali e quali valori categorici.
3. Visualizzare con un opportuno grafico i contenuti dell'attributo *Anno*, commentando in forma scritta il risultato della visualizzazione, in particolare riconoscendo un modello probabilistico per la descrizione del fenomeno.
4. Calcolare la tabella delle frequenze relative del numero di feriti.
5. Visualizzare tale tabella attraverso un grafico, in modo che sia facile vedere qual è (approssimativamente) la probabilità che, a seguito di un incidente, vi sia almeno un ferito.
6. Fornire una stima della probabilità che, a seguito di un incidente, vi sia almeno un ferito.
7. Fornire una stima del numero atteso di feriti a seguito di un incidente.
8. Quale tra le famiglie di funzioni massa di probabilità studiate appare più indicata per descrivere la distribuzione del numero di feriti? Si giustifichi la risposta.
9. Fornire una stima del parametro di tale distribuzione
10. Visualizzare graficamente la legge di probabilità del modello teorico individuato.
11. Visualizzare il diagramma di dispersione che mette in relazione il numero di feriti con il numero di illesi. Come è possibile interpretare il risultato ottenuto in modo qualitativo?
12. Si giustifichi in termini quantitativi l'interpretazione data al punto precedente (cioè calcolando un opportuno indicatore statistico).

Esercizio II

1. Indicato con X_1, \dots, X_n un campione casuale di taglia n estratto dalla popolazione $X =$ numero di illesi a seguito di un incidente, si proponga uno stimatore T_n per il valore atteso μ di X .
2. Si esprima la deviazione standard di T_n in funzione di n e della varianza di X .
Indichiamo con Z una variabile aleatoria normale standard.
3. Si dimostri che $P(|Z| < k) = 2\Phi(k) - 1$, dove Φ denota la funzione di distribuzione cumulativa di Z .

4. Determinare il valore k tale che $P(|Z| < k) = 0.99$.
5. Utilizzando il campione a disposizione e un'opportuna approssimazione della distribuzione di T_n , si determini una condizione sufficiente su ϵ in modo che l'errore in valore assoluto compiuto stimando μ con T_n sia minore di ϵ con probabilità 0.99.