Esercizi di Statistica della 4^a settimana (Corso di Laurea in Biotecnologie, Università degli Studi di Padova).

Esercizio 1. Il cosiddetto "test del DNA" non fa altro che misurare la lunghezza di K geni, senza controllare le basi azotate che li compongono. Per ognuno di tali geni, la probabilità che due dati individui presentino una lunghezza uguale viene assunta come pari a $1/10$. Un'altra ipotesi che viene comunemente fatta è che le lunghezze di geni diversi siano indipendenti l'una dall'altra.
Supponiamo di misurare la lunghezza di $K = 7$ geni da un campione di DNA trovato su una "scena del crimine".

1. Calcolare la probabilità che un dato individuo abbia la lunghezza dei suoi geni uguali a quella del campione.
2. Supponendo di avere un database di $n = 101905$ individui, calcolare la probabilità che almeno uno di questi individui abbia le lunghezze dei suoi geni uguali a quelle del campione incriminato.
3. Calcolare la probabilità che almeno due di questi individui abbiano le lunghezze dei loro geni uguali a quelle del campione incriminato.
4. Supponendo di aver trovato un individuo con le lunghezze dei geni uguali a quelle del campione incriminato, calcolare la probabilità che ce ne sia almeno un altro.

Esercizio 2. Sia $X \sim B(30; 1/3)$ e supponiamo di voler calcolare le probabilità $\mathbb{P}\{X \leq 15\}$ e $\mathbb{P}\{X \geq 16\}$.

1. Dire se si può utilizzare l'approssimazione normale, motivando la risposta.
2. Calcolare $\mathbb{P}\{X \leq 15\}$.
3. Calcolare $\mathbb{P}\{X \geq 16\}$ e confrontare il risultato con quanto ottenuto al punto 2.
4. Come si fa a far venire uguali i due risultati sopra?

Esercizio 3. Si supponga che il peso (in tonnellate) di un autoveicolo si distribuisca come una variabile aleatoria di media 3 e deviazione standard 0.3. Nel seguito, supporremo di poter applicare l'approssimazione normale.

1. Se consideriamo n autoveicoli con "n grande", con che variabile aleatoria possiamo approssimare il peso totale?
2. Supponiamo che la portata della campata di un ponte sia 400 tonnellate, prima di riportare danni strutturali. Se il numero massimo di veicoli che ci possono transitare contemporaneamente è uguale a 100, qual è la probabilità che si possa danneggiare?
3. Rispondere alla stessa domanda supponendo che la portata della campata sia una variabile aleatoria gaussiana di media 400 e di deviazione standard 40.

4. Supponiamo di voler controllare se l'assunzione iniziale (media = 3, dev. standard = 0.3) era corretta. Pesando 10 autoveicoli, otteniamo i seguenti valori (in tonnellate):

$$2.53 \quad 1.91 \quad 3.01 \quad 6.12 \quad 3.42 \quad 2.95 \quad 3.24 \quad 2.89 \quad 4.52 \quad 3.07$$

Quali sono la media e la varianza stimate da questo campione?

Esercizio 4. Supponiamo di voler studiare la relazione tra abuso di analgesici e livello di creatinina nel sangue. In particolare, consideriamo 15 persone che lavorano in una fabbrica e sono conosciuti per "abuso di analgesici" (cioè più di 10 pillole al giorno) e misuriamo il loro livello di creatinina, con i seguenti risultati:

$$0.9, \quad 1.1, \quad 1.6, \quad 2.0, \quad 0.8, \quad 0.7, \quad 1.4, \quad 1.2, \quad 1.5, \quad 0.8, \quad 1.0, \quad 1.1, \quad 1.4, \quad 2.2, \quad 1.4$$

1. Stimare media, deviazione standard ed errore standard della media del livello di creatinina in base ai dati sopra.
2. Supponendo di sapere che la vera deviazione standard è uguale a quella stimata, calcolare l'intervallo di confidenza al 95% della media.
3. Supponendo sempre di conoscere la vera deviazione standard, quanto grande dovremo prendere n in modo che l'intervallo di confidenza sia largo meno di 0.1?
4. Supponiamo ora di sapere che nella popolazione "normale" il livello di creatinina sia 1.0. Dire, in base all'intervallo di confidenza, se si ritiene plausibile che il livello medio sia 1.0 anche in questo campione.

Esercizio 5. L'ente che gestisce un tratto di autostrada conserva sale a sufficienza per eliminare un tratto di 80 pollici di neve. Supponiamo che la quantità di neve che cade al giorno sia una variabile aleatoria di media 1.5 pollici e deviazione standard di 0.3 pollici.

1. Trova la probabilità approssimata che il sale a disposizione basti per 50 giorni.
2. Quali sono le ipotesi fatte per rispondere al punto 1.? Possono ritenersi giustificate?
3. Supponiamo che nei primi 10 giorni del periodo siano caduti in totale 20 pollici di neve. Supponendo che la deviazione standard sia sempre pari a 0.3, calcolare l'intervallo di confidenza al 99% per la media.
4. Possiamo ancora supporre che la media sia 1.5? I calcoli del punto 1. sono quindi ancora giustificati? (non eseguire nuovi calcoli)

Soluzioni su `http://www.math.unipd.it/~vargiolu/Statistica/`