Compito scritto dell'esame di Statistica e analisi dei dati 06 settembre 2021	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e' di 2 ore. Durante la prova e' possibile consultare il formulario ed utilizzare la calcolatrice. Non e' possibile consultare libri e appunti.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu' passaggi di calcolo, non sara' preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu' frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara' eventualmente rappresentato in quest'ultima forma).

Problemi

Esercizio 1.

Quando esce di casa al mattino Aldo non prende mai l'ombrello se il cielo é sereno, lo prende sempre se il cielo é nuvoloso e infine, quando il tempo é incerto, decide se prendere o no l'ombrello lanciando una moneta regolare. Secondo i dati storici, nella cittá dove vive Aldo il tempo é sereno (= evento H_1), nuvoloso (= evento H_2) o incerto (= evento H_3) nelle proporzioni $10 : 3 : 2$, rispettivamente.

(a) Nell'arco di un anno (= 365 giorni) quante volte all'incirca Aldo esce con l'ombrello?

Soluzione: Definito l'evento

$$A = \text{Aldo esce con l'ombrello}$$

i dati sono $P(H_1) = 10/15$, $P(H_2) = 3/15$, $P(H_3) = 2/15$, $P(A \mid H_1) = 0$, $P(A \mid H_2) = 1$, P (A | H_3) = 1/2. Allora

$$P(A) = \sum_i P(A \mid H_i)P(H_i) = \frac{4}{15} \simeq 0.27$$

Quindi Aldo esce con l'ombrello mediamente

$$365 \times \frac{4}{15} \simeq 97$$

giorni all'anno.

(b) Se il 30 marzo Aldo é uscito con l'ombrello, qual é la probabilitá che quel giorno fosse nuvoloso?

Soluzione: Se il 30 marzo Aldo é uscito con l'ombrello, la probabilitá che quel giorno fosse nuvoloso vale

$$P(H_2 \mid A) = \frac{P(A \mid H_2)P(H_2)}{P(A)} = \frac{3/15}{4/15} = \frac{3}{4} = 0.75$$

Esercizio 2.

Supponiamo che il 10% dei semi di una certa varietá di pianta siano sterili.

(a) Qual é la probabilitá che germini almeno il 95% dei semi contenuti in una confezione da 20 semi (= evento A)?

Soluzione: Detto X_n il numero di semi sterili in una confezione di n semi, si puó assumere che sia $X_n \sim Bin(n; 0.1)$. Si chiede la probabilitá che una confezione di 20semi contenga al massimo $20 \times 5\% = 1$ seme sterile

$$P(A) = P(X_{20} \leq 1) = \sum_{x=0}^{1} \binom{20}{x} (0.1)^x (0.9)^{20-x} \approx 39.17\%$$

(b) Qual é la probabilitá che germini almeno il 95% dei semi contenuti in una confezione da 100 semi (= evento B)?

Soluzione: (Distribuzione binomiale in approssimazione poissoniana o Gaussiana) Analogamente alla precedente, si chiede la probabilitá che una confezione di 100semi contenga al massimo $100 \times 5\% = 5$ semi sterili, dunque

$$P(B) = P(X_{100} \leq 5) = \sum_{x=0}^{5} \binom{100}{x} (0.1)^x (0.9)^{100-x}$$

Possiamo in questo caso approssimare sia con Poisson (piccola probabilitá $p = 0.1$) sia con la Normale ($n = 100$)

$$Bin(100; 0.1) \approx Pois(10)$$

$$Bin(100; 0.1) \approx \mathcal{N}(10; 3^2)$$

Con Poisson si ha che

$$P(B) = P(X_{100} \le 5) \approx \sum_{x=0}^{5} \frac{10^x}{x!} e^{-10} \simeq 6.71\%.$$

Con la normale usando la correzione di continuitá e usando le tabelle della normale ridotta:

$$P(B) = P(X_{100} \le 5) \approx F\left(\frac{5.5 - 10}{3}\right) \simeq 6.68\%.$$

(si osserva che l'approssimazione normale é lievemente migliore: calcolando esattamente con la Binomiale $P(B) \simeq 5.76\%$)

Esercizio 3. Tra le 23 e le 23:30 lungo un tratto di strada transitano in media 40 veicoli.

(a) Qual è la probabilità di osservare, tra un veicolo e il successivo, un intervallo di tempo minore di 10 sec e uno piú lungo di 5 min?

Soluzione: E' un problema di tempi d'attesa modellabile con la distribuzione esponenziale negativa.
Su 30 min il tempo medio (in secondi) tra due veicoli successivi é $\frac{30\times 60}{40} = 45sec$ a cui corrisponde un rate di arrivo pari a

$$\lambda = \frac{1}{45} = 0.022 sec^{-1}$$

Per $t = 10$,

$$P(0 \le T \le 10) = F(10) = 1 - e^{-0.022\cdot 10} = 0.199 \approx 20\%$$

Intervalli piú lunghi di $t = 5min = 300sec$ saranno osservati con probabilitá

$$P(T > 300) = 1 - P(0 \le T \le 300) = e^{-0.022\cdot 300} = 1.27 \times 10^{-3} \approx 0.13\%$$

Esercizio 4. Un segnale di intensitá media incognita, trasmesso dalla sorgente S, é raccolto dal ricevitore R con un rumore additivo gaussiano di media nulla e deviazione standard uguale a 2 (pertanto il segnale avrá la stessa distribuzione del rumore, con uguale deviazione standard ma media incognita). Per ridurre l'errore, lo stesso segnale viene inviato 10 volte da S ad R e la media campionaria dei segnali ricevuti é 8.7.

(a) Qual é la confidenza che il segnale trasmesso sia compreso fra 7.0 e 10.4?

Soluzione:
Il segnale X raccolto al ricevitore é variabile aleatoria con distribuzione normale $X \sim \mathcal{N}(\mu, \sigma^2) = \mathcal{N}(\mu, 2^2)$, dunque con μ incognita ma $\sigma = 2$ nota.
Il problema richiede di stimare il livello di confidenza $1-\alpha$ tale che l' IC a due code [7.0,10.4] per la media μ soddisfi

$$P(7.0 < \mu < 10.4) = 1 - \alpha.$$

Poiché X proviene da popolazione normale con μ incognita e σ nota, la stima intervallare, sapendo che $n = 10$ e $\bar{x} = 8.7$, si scrive come

$$P\left(\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

ovvero $\mu = \bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ con confidenza $1 - \alpha$.
In questo caso $z_{\frac{\alpha}{2}}$ é incognito, perché non conosciamo il livello di fiducia α, e va determinato. Assumendo $\mu = \bar{x}$, il semi intervallo di confidenza (simmetrico) ha lunghezza

$$10.4 - \bar{x} = 1.7 = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

da cui $z_{\frac{\alpha}{2}} = \frac{1.7\cdot\sqrt{10}}{2} \approx 2.7$
Dalla tabella normale $F(z_{\frac{\alpha}{2}}) \approx F(2.7) = 0.9965 = 1 - \frac{\alpha}{2}$.
Per le proprietá della cumulativa standard

$$1 - \alpha = P(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}) = 2F(z_{\frac{\alpha}{2}}) - 1 \approx 2 \cdot 0.9965 - 1 = 0.993 = 99.3\%$$

che rappresenta il livello di confidenza dell'IC a due code [7.0,10.4].

Esercizio 5. Il termostato di un condizionatore d'aria é supposto essere tarato sul valore di soglia $\mu = 25°C$ (cioé l'apparecchio entra in funzione quando la temperatura dell'ambiente supera i $25°C$). Un tecnico misura in 8 occasioni diverse la temperatura X alla quale il termostato scatta, ottenendo i valori seguenti ($°C$):

24.6 24.8 25.2 25.4 25.5 24.0 24.7 25.3

(a) Formulando un adeguato modello per X, stabilire se, al 95% di confidenza, il valore di μ sia plausibile statisticamente

Soluzione:
Assumendo una distribuzione normale della temperatura, $X \sim \mathcal{N}(\mu, \sigma^2)$, poiché la varianza σ^2 é incognita, la stima intervallare di μ al 95% di confidenza si ottiene considerando

$$P\left(\bar{x} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right) = 1 - \alpha = 0.95$$

Poiché al 95% di confidenza $\alpha = 0.05$, allora $t_{\frac{\alpha}{2}} = t_{0.025}$, da determinarsi dalla distribuzione di Student con $\nu = n - 1 = 7$ gdl:

$$t_{0.025} = 2.365$$

Dai dati conosciuti si ottengono media e deviazione standard campionaria:

$$\bar{x} \approx 24.93$$

$$s \approx 0.83$$

Pertanto

$$24.93 - 2.365 \times \frac{0.83}{2.64} < \mu < 24.93 + 2.365 \times \frac{0.83}{2.64}$$

ovvero $24.2 < \mu < 25.66$ e dunque il valore $\mu = 25°C$ puó essere ritenuto plausibile.

(b) Stabilire quanto vale al massimo la precisione p (ovvero il reciproco della deviazione standard) del termostato al 95% di confidenza

Soluzione:
Assumendo $\mu = 25°C$ per l'analisi precedente, $X \sim \mathcal{N}(25, \sigma^2)$.
Sappiamo che $p = \frac{1}{\sigma}$, dunque il valore di precisione massima é quello di deviazione standard minima, $p_{max} = \frac{1}{\sigma_{min}}$.
Il problema richiede di determinare

$$P(p < p_{max}) = 1 - \alpha = 0.95$$

Dalla definizione di p questo equivale alla stima unilaterale sul limite inferiore della varianza (σ^2_{min}) :

$$P(p < p_{max}) = P\left(\sqrt{\frac{(n-1)s^2}{\chi^2_\alpha}} < \sigma\right) = 1 - \alpha$$

Ne deriva che al 95% di confidenza il limite superiore di precisione vale $p_{max} = \sqrt{\frac{\chi^2_{0.05}}{(n-1)s^2}}$, dunque, leggendo dalla tabella della distribuzione Chi-quadro, per $\nu = 7$, $\chi^2_{0.05}(7) = 14.067$ e usando s^2 calcolato precedentemente.

$$p < \sqrt{\frac{\chi^2_{0.05}}{(n-1)s^2}} \approx 1.69°C$$