Esame Scritto del Sesto Appello

Tempo a disposizione: 2 ore

Riportare la matricola **all'inizio** di ogni foglio. La soluzione di ogni esercizio deve iniziare in una nuova pagina ed essere scritta in modo chiaro e ordinato: non verrà valutata se la calligrafia è illeggibile. Si riporti il procedimento dettagliato che porta alle risposte, risposte corrette non adeguatamente motivate saranno penalizzate.
Non è permesso l'uso di note, appunti, manuali o materiale didattico di alcun tipo. Non è permesso l'uso di dispositivi elettronici ad esclusiva eccezione di una calcolatrice non programmabile.

1. Per ognuna delle seguenti affermazioni si determini se essa è VERA oppure FALSA, motivando rigorosamente le risposte.

   (a) Se due variabili aleatorie X, Y hanno la stessa distribuzione di probabilità, allora non sono indipendenti.
   FALSO: le singole componenti di una scelta uniforme su $\{(0,0),(0,1),(1,0),(1,1)\}$ sono Bernoulli($\frac{1}{2}$) indipendenti.

   (b) Due variabili aleatorie X, Y soddisfano una relazione lineare $Y = aX + b$ se e solo se la loro correlazione vale $\rho(X,Y) = \pm 1$.
   VERO: come visto nel corso vale

   $$\min_{a,b} \mathbb{E}[(Y - aX - b)^2] = \mathrm{Var}(Y)(1 - \rho(X,Y)^2),$$

   per cui quando il minimo a sinistra è nullo deve valere $\rho(X,Y) = 1$.

   (c) Data una generica variabile aleatoria X, la sua funzione di ripartizione è data dall'integrale $F_X(t) = \int_0^t f(x)dx$ di una certa funzione f.
   FALSO: ciò è vero —per definizione— solo per le variabili con densità, e per cui la densità f sia zero per $x < 0$; ad esempio non esiste una funzione f che soddisfa la relazione proposta se X è una variabile di Bernoulli.

   (d) Scambiare l'ipotesi nulla e quella alternativa può cambiare l'esito di un test statistico.
   VERO: le condizioni su livello e potenza non sono simmetriche nelle due ipotesi. Ad esempio se si è misurata l'altezza di 100 persone ottenendo media campionaria 175.5cm, supponendo deviazione standard 5cm, consideriamo due test con ipotesi nulle $H_0) m \leq 175$ e $H_0') m > 175$. In entrambi i casi $Z = \sqrt{N}(\bar{X} - m)/\sigma = 1$, e i quantili da confrontare sono $q_{0.95} = 1.64$ e $q_{0.05} = -1.64$. In particolare, in entrambi i casi siamo nella regione critica, quindi scambiare un ipotesi nulla che sta venendo rigettata con la corrispondente alternativa non conduce ad accettazione.

   (e) In generale, più basso è il livello imposto a un test statistico, inferiore è la potenza.
   VERO: il livello è la probabilità sotto H_0 della regione critica, ed imporlo più basso significa ridurre la regione critica; a sua volta la potenza è la probabilità sotto l'ipotesi alternativa della regione critica, ma se quest'ultima si è ridotta la potenza non può aumentare.

   (f) In un test statistico una forte evidenza a favore dell'ipotesi nulla può produrre un p-value superiore a 1.
   FALSO: il p-value è una probabilità e non può mai assumere valori superiori a 1.

2. Il numero di volte in cui un individuo contrae l'influenza (in un anno) ha distribuzione di Poisson di parametro $\lambda = 0.3$.

   (a) Qual è la probabilità che un individuo contragga l'influenza almeno due volte (in un anno)?
   Chiamiamo X la v.a. che conta il numero di volte in cui l'individuo contrae l'influenza. La probabilità cercata è

   $$\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) = 1 - (1 + \lambda)e^{-\lambda} = 0.0369.$$

   Per una persona vaccinata contro l'influenza, se il vaccino è efficace, il numero di volte (in un anno) in cui questa persona contrae l'influenza ha distribuzione di Poisson di parametro $\lambda/2 = 0.15$ (cioè il vaccino, quando efficace, dimezza l'incidenza dell'influenza); se invece il vaccino non è efficace, la distribuzione rimane di Poisson di parametro $\lambda = 0.3$. Supponiamo che il vaccino sia efficace nel 75% delle persone vaccinate.

   (b) Per una persona vaccinata, qual è la probabilità di contrarre l'influenza almeno due volte?

   Indichiamo con E ed E^c gli eventi "il vaccino è stato efficace" e il suo complementare, e indichiamo con Y il numero di volte in cui un individuo vaccinato contrae l'influenza. Per il punto precedente, la probabilità di contrarre l'influenza dato E^c è

   $$\mathbb{P}(Y \geq 2 \mid E^c) = 0.0369.$$

   Se il vaccino è efficace, la v.a. Y ha distribuzione di Poisson di parametro $\lambda/2$ e quindi la probabilità di contrarre l'influenza dato E è

   $$\mathbb{P}(Y \geq 2 \mid E) = 1 - (1 + \lambda/2)e^{-\lambda/2} = 0.0102.$$

   Per ipotesi $P(E) = 0.75$. Per la formula della probabilità totale, la probabilità cercata è

   $$\mathbb{P}(Y \geq 2) = \mathbb{P}(Y \geq 2 \mid E)\mathbb{P}(E) + \mathbb{P}(Y \geq 2 \mid E^c)\mathbb{P}(E^c) = 0.0169.$$

   (c) Se una persona contrae l'influenza almeno due volte, qual è la probabilità che il vaccino sia stato efficace su questa persona?
   Per la formula di Bayes, la probabilità cercata è

   $$\mathbb{P}(E \mid Y \geq 2) = \frac{\mathbb{P}(Y \geq 2 \mid E)\mathbb{P}(E)}{\mathbb{P}(Y \geq 2)} = 0.4527.$$

3. Il tempo di elaborazione (espresso in secondi) di un certo processo, da parte di un dato server, ha distribuzione gaussiana di valore atteso incognito e deviazione standard 1 secondo. L'azienda produttrice del server dichiara un valore atteso di 5 secondi. Su 100 prove effettuate, risulta un tempo medio di 5.3 secondi.

(a) Sulla base dei dati, fornire un intervallo di fiducia di livello 95% per il valore atteso del tempo di elaborazione.
Stiamo cercando un intervallo di fiducia per la media di una popolazione gaussiana, deviazione standard nota $\sigma = 1$. L'invervallo di fiducia cercato è ($n = 100$, $\alpha = 0.05$)

$$\left[\bar{X} \pm \frac{\sigma}{\sqrt{n}} q_{1-\alpha/2}\right] = \left[\bar{X} \pm 0.196\right].$$

Inserendo il valore $\bar{x} = 5.3$, troviamo $[5.104, 5.496]$.

(b) L'affermazione dell'azienda è plausibile? Formulare un opportuno test di ipotesi a livello 5% ed applicarlo ai dati in esame.
Siamo in presenza di un test di ipotesi sulla media m di una popolazione gaussiana, con deviazione standard $\sigma = 1$ nota. L'ipotesi nulla è $H_0 : m = 5 (= m_0)$, contro $H_1 : m \neq 5$. La regione critica di livello $\alpha = 0.05$ è ($n = 100$)

$$C = \{\frac{\sqrt{n}}{\sigma}|\bar{X} - m_0| > q_{1-\alpha/2}\} = \{|\bar{X} - 5| > 0.196\}$$

Applichiamo il test al dato $\bar{x} = 5.3$: i dati cadono nella regione critica C, quindi c'è evidenza, a livello 0.05, contro l'affermazione dell'azienda (valore atteso pari a 5).
Poiché il test è bilatero, alla stessa conclusione si può giungere semplicemente osservando che 5.3 non cade nell'intervallo di fiducia trovato al punto precedente.
In alternativa, si può calcolare il p-value relativo a $\bar{x} = 5.3$, che è (chiamando Z una v.a. normale standard)

$$\bar{\alpha} = P\left(|Z| > \frac{\sqrt{n}}{\sigma}(\bar{x} - m_0)\right) = 2 \cdot (1 - \Phi(\bar{x} - m_0)) = 2 \cdot (1 - \Phi(3)) = 0.0026 < 0.05.$$

Un nuovo server viene immesso sul mercato, per il quale l'azienda dichiara un tempo medio di elaborazione di 4 secondi. La distribuzione (del tempo di elaborazione) rimane gaussiana con deviazione standard 1 secondo. Su questo server non sono noti molti dati, ma si sa che, in (esattamente) 30 prove su 100, il server ha impiegato più di 5 secondi.

(a) Sulla base di questo dato (tempo medio maggiore di 5 secondi in 30 prove su 100), l'affermazione dell'azienda sul nuovo server è plausibile? Formulare un opportuno test di ipotesi a livello 5% ed applicarlo al dato in esame.
Assumendo l'affermazione dell'azienda, il tempo di elaborazione del nuovo server ha distribuzione $Y \sim N(5, 1)$ e quindi la probabilità p che il nuovo server impieghi più di 5 secondi è

$$p_0 := \mathbb{P}(Y > 5) = P(Z > 1) = 1 - \Phi(1) = 0.1587.$$

Verifichiamo quindi se un tale valore p_0 è plausibile o meno, con un test su una proporzione per campione Bernoulli di grande numerosità. L'ipotesi H_0 è $p = p_0 = 0.1587$ contro $H_1 : p \neq 0.1587$. Detta $\bar{X}$ la frequenza relativa campionaria (dell'evento "più di 6 secondi"), la regione critica è ($n = 100$)

$$C = \{\sqrt{n}\frac{|\bar{X} - p_0|}{\sqrt{p_0(1 - p_0)}} > q_{1-\alpha/2}\} = \{10\frac{|\bar{X} - 0.1587|}{\sqrt{0.8413 \cdot 0.1587}} > 1.96\}$$

Il valore $\bar{x} = 30/100 = 0.3$ cade nella regione critica, quindi l'affermazione dell'azienda non è plausibile a livello 0.05.
In alternativa, si può anche calcolare il p-value:

$$\bar{\alpha} = 2 \cdot (1 - \Phi(10 \frac{\bar{x} - 0.1587}{\sqrt{0.8413 \cdot 0.1587}})) = 2 \cdot (1 - \Phi(3.87)) \approx 0,$$

quindi l'affermazione dell'azienda non è plausibile a ogni ragionevole livello.

Valori numerici utilizzabili:

$$\Phi(1) = 0.8413, \quad \Phi(2) = 0.9773, \quad \Phi(3) = 0.9987,$$
$$q_{0.95} = 1.64, \quad q_{0.975} = 1.96, \quad q_{0.99} = 2.33, \quad q_{0.995} = 2.58,$$
$$t_{0.95,99} = 1.66, \quad t_{0.975,99} = 1.98, \quad t_{0.99,99} = 2.36, \quad t_{0.995,99} = 2.63.$$