Esame Scritto del Primo Appello

Tempo a disposizione: 2 ore

Riportare il numero di matricola **all'inizio** di ogni foglio. La soluzione di ogni esercizio deve essere scritta in modo chiaro e ordinato, e può non essere valutata se la calligrafia è illeggibile. Se l'esercizio lo richiede, evidenziare il risultato numerico nella soluzione. Le soluzioni degli esercizi devono essere riportate sul foglio protocollo nell'ordine proposto, la soluzione di ogni esercizio deve iniziare in una nuova pagina.
Le soluzioni devono includere 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, al di fuori del formulario e delle tavole statistiche fornite assieme al compito. Non è permesso l'uso di dispositivi elettronici ad esclusiva eccezione di una calcolatrice non programmabile. L'infrazione di queste regole o la comunicazione con altri comportano l'annullamento del compito.
In ROSSO si riportano alcuni errori comuni riscontrati nella correzione.

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

   (a) Una variabile aleatoria $X : \Omega \to \mathbb{R}$ costante, ovvero $X(\omega) = c$ per ogni $\omega \in \Omega$, è indipendente da qualsiasi altra variabile aleatoria $Y : \Omega \to \mathbb{R}$.
   VERO: l'evento $\{X \in A\}$ è sempre uno tra $\emptyset$ e Ω, ma questi ultimi sono indipendenti da ogni evento $\{Y \in B\}$.
   RISPOSTE ERRATE: non è sufficiente affermare che "l'informazione su X non influenza quella su altre variabili", ma va verificata la definizione.

   (b) Il β-quantile di una variabile aleatoria con densità esponenziale di parametro $\lambda > 0$ è $-\log(1-\beta)/\lambda$.
   VERO: $F(x) = 1 - e^{-\lambda x} = \beta$ se e solo se $x = -\log(1-\beta)/\lambda$.

   (c) Due variabili aleatorie scorrelate sono anche indipendenti.
   FALSO: esempio: X uniforme su $[-1, 1]$ e X^2 sono scorrelate ma non indipendenti.
   RISPOSTE ERRATE: l'esempio riportato è stato visto in classe, ma è imprescindibile scegliere la legge di X simmetrica rispetto a 0, altrimenti l'esempio non funziona.

   (d) La media campionaria $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$ di un campione di variabili aleatorie con momento secondo finito è uno stimatore corretto e consistente del parametro $m = \mathbb{E}[X_i]$.
   VERO: la correttezza discende dalla linearità del valore atteso, mentre la consistenza dalla Legge Debole dei Grandi Numeri.
   RISPOSTE ERRATE: alcuni candidati hanno negato che siano soddisfatte le ipotesi della Legge dei Grandi Numeri, ma va ricordato che le variabili sono i.i.d. per definizione di campione statistico, e che possedendo momento secondo hanno anche valore atteso finito.

   (e) Un intervallo di fiducia ha livello 95% se esso contiene almeno il 95% dei dati del campione.
   FALSO: non ci sono legami tra intervallo di fiducia e numero di dati in esso contenuti.

   (f) Se X è una v.a. discreta con funzione di massa p_X, la funzione di ripartizione F_X di X è $F_X(x) = \int_{-\infty}^{x} p_X(y)dy$.
   FALSO: La funzione di ripartizione è $F_X(x) = \sum_{x_i \le x} p_X(x_i)$.

RISPOSTE ERRATE: dei candidati hanno affermato che la funzione di ripartizione non è definita per variabili discrete, e ciò è del tutto FALSO.

2. Sia Φ la funzione di ripartizione della variabile Gaussiana standard e sia $\mu > 0$ un parametro.

   (a) Si mostri che la funzione F_μ definita da

$$F_\mu(t) = \begin{cases} 2 - 2\Phi\left(\frac{\mu}{\sqrt{t}}\right) & \text{per } t > 0 \\ 0 & \text{per } t \leq 0 \end{cases}$$

   è la funzione di ripartizione di una variabile aleatoria.
   Ricordando che $\Phi(0) = 1/2$, che Φ è C^1 e $\Phi'(x) = \varphi(x)$ densità gaussiana standard, abbiamo

$$\lim_{t\to+\infty} F(t) = \lim_{x\to 0+} (2 - 2\Phi(x)) = 2 - 2\Phi(0) = 1,$$
$$\lim_{t\to 0+} F(t) = \lim_{x\to+\infty} (2 - 2\Phi(x)) = 0 = F(0),$$
$$\lim_{t\to-\infty} F(t) = 0$$

   in particolare F è continua a destra (in realtà continua) su $\mathbb{R}$ e ha i limiti giusti a $\pm\infty$. Inoltre F è costante su $(-\infty, 0]$, continua in $t = 0$ e non-decrescente su $(0, +\infty)$, come si vede derivando F (o notando che F è composizione di due funzioni, $2 - 2\Phi$ e $\mu/\sqrt{t}$, entrambe non-crescenti); quindi F è non-decrescente. Dunque F è una funzione di ripartizione.

   (b) Si mostri che una variabile aleatoria X con funzione di ripartizione F_μ ha densità

$$f_\mu(t) = \frac{\mu}{\sqrt{2\pi t^3}} \exp\left(-\frac{\mu^2}{2t}\right) 1_{(0,+\infty)}(t).$$

   F è continua su $\mathbb{R}$ e differenziabile sui tratti $(-\infty, 0)$ e $(0, +\infty)$. Quindi X ammette densità

$$f(t) = F'(t) = -2\Phi'(\mu/\sqrt{t}) \frac{\mu}{-2\sqrt{t^3}} 1_{(0,+\infty)}(t) = \frac{1}{\sqrt{t^3}} \varphi(\mu/\sqrt{t}) 1_{(0,+\infty)}(t).$$

   (c) Si consideri un campione di variabili aleatorie $X_1, \ldots, X_n$ i.i.d. con funzione di ripartizione F_μ (dipendente dal parametro $\mu > 0$); si determini lo stimatore di massima verosimiglianza per μ.
   Dati esiti $x_1, \ldots, x_n > 0$, la verosimiglianza è

$$L(\mu; x_1, \ldots, x_n) = (2\pi)^{-n/2} (x_1 \cdots x_n)^{-3/2} \mu^n \exp\left(-\mu^2 \sum_{i=1}^n \frac{1}{2x_i}\right).$$

   Derivando, troviamo

$$\frac{d}{d\mu} L(\mu; x_1, \ldots x_n) = 0 \Leftrightarrow \hat{\mu} = \frac{1}{\sqrt{\frac{1}{n}\sum_{i=1}^n \frac{1}{x_i}}}.$$

Inoltre i limiti di $L(\mu; x_1, \dots x_n)$ per $\mu \to 0$ e per $\mu \to +\infty$ sono entrambi nulli. Perciò

$$\hat{\mu} = \frac{1}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}}}$$

è l'unico punto di massimo di L, quindi l'unico stimatore di massima verosimiglianza.

3. Una società vuole testare l'uso di un software per ridurre la temperatura in esercizio di un modello di PC. Il software viene testato su 100 PC, su cui viene rilevata la temperatura prima e dopo l'installazione del software. I dati forniscono una temperatura media di 72 gradi (Celsius) prima dell'installazione e di 63 gradi dopo l'installazione, con deviazioni standard campionarie di 6 e di 7 gradi rispettivamente prima e dopo l'installazione; la deviazione standard campionaria della differenza di temperatura prima e dopo risulta di 5 gradi.

   (a) C'è evidenza che la temperatura media dopo l'installazione sia inferiore ai 64 gradi? Formulare ed applicare un opportuno test di ipotesi e calcolarne il p-value.
   Siamo nel caso di test T sulla media di un campione grande (varianza non nota) Osserviamo che per la numerosità $n = 100$ la cdf e i quantili della distribuzione di Student con $n-1$ gradi di libertà sono approssimabili con quelli della Gaussiana standard, dunque usiamo questi ultimi nei calcoli che seguono.
   Sia X la v.a. che indica la temperatura dopo l'installazione del software, $m = E[X]$. Testiamo $H_0 : m \geq 64$ contro $H_1 : m < 64$. La statistica di test (con distribuzione approssimativamente $N(0,1)$ per $m = m_0$) e la regione critica sono

$$Z = \frac{\sqrt{n}}{S_n}(\bar{X}_n - m_0) = \frac{10}{S_n}(\bar{X}_n - 64),$$
$$C = \{Z < q_\alpha = -q_{1-\alpha}\}.$$

   Il p-value dei dati è, essendo $z = -1.43$ il valore assunto dalla statistica test,

$$\bar{\alpha} = P\{Z < z\} = \Phi(-1.43) = 1 - \Phi(1.43) = 0.076$$

   che conduce ad accettare l'ipotesi nulla fino a livello 7.6%, ossia l'ipotesi nulla è mediamente plausibile.

   (b) C'è evidenza che, mediamente, la temperatura diminuisca con l'installazione del software? Formulare ed applicare un opportuno test di ipotesi e calcolarne il p-value.
   Effettuiamo un T-test di confronto tra medie di campioni accoppiati, con varianza della differenza non nota: poichè $n = 100$ come detto sopra usiamo cdf e quantili della Gaussiana standard invece che della distribuzione Student $T(n-1)$.
   Sia X la differenza tra temperatura dopo e prima dell'installazione del software, $m = E[X]$. Testiamo $H_0 : m \geq 0$ contro $H_1 : m < 0$. La statistica di test (con distribuzione approssimativamente $N(0,1)$ per $m = m_0$) e la regione critica sono

$$Z = \frac{\sqrt{n}}{S_n}(\bar{X}_n - m_0) = \frac{10}{S_n}\bar{X}_n,$$
$$C = \{Z < q_\alpha = -q_{1-\alpha}\}.$$

   Il p-value dei dati è, essendo $z = 2 \cdot (-9) = -18$ il valore assunto dalla statistica test,

$$\bar{\alpha} = P\{Z < z\} = \Phi(-18) \approx 0.$$

Quindi c'è evidenza, ad ogni ragionevole livello, di una diminuzione media delle temperature, ovvero l'ipotesi nulla non è plausibile.
RISPOSTE ERRATE: non va confuso il test per campioni accoppiati (da applicare in questo caso) con quello per campioni indipendenti (ipotesi NON verificata in questo caso).

(c) Fornire un intervallo di fiducia di livello 95% per la diminuzione media della temperatura.

Dobbiamo determinare un intervallo di fiducia per un campione di taglia grande $n = 100$ con varianza non nota, e ancora una volta approssimiamo i quantili di Student con $n-1$ gradi di libertà con quelli Gaussiani standard. L'intervallo cercato è

$$[\bar{X}_n \pm \frac{S_n}{\sqrt{n}} q_{1-\alpha/2}] = [\bar{X}_n \pm \frac{S_n}{10} \cdot 1.96]$$

e il suo valore numerico è $[9 \pm 0.98]$.