Inferenza Statistica

Esame del 11 febbraio 2013

Tempo a disposizione 2 ore.

Tra parentesi quadre i punteggi massimi attribuibili per ciascun quesito (Totale: 36).

1. Vi sono numerosi studi il cui obiettivo è quello di comparare il comportamento di consumo maschile e femminile. Si vuole valutare se ci siano differenze nella spesa mensile (in euro) media per abbigliamento di uomini e donne. Si assuma in prima istanza che sia nota la varianza nelle due popolazioni. I risultati di un recente studio sono riportati sotto:

	Uomini	donne
dimensione campionaria	34	28
media campionaria	212	256
varianza (nota) delle due popolazioni	1656	1540

Si vuole sottoporre a verifica l'ipotesi di uguaglianza della spesa media per le due popolazioni contro l'alternativa che sia maggiore quella delle femmine.

a. [3.] Se si assume la normalità delle due popolazioni, qual è la funzione test opportuna e come è distribuita tale funzione test sotto H_0? Ottenere il p-value per verificare l'ipotesi sopra data e decidere riguardo ad essa. Ottenere inoltre la regione di rifiuto per $\alpha = 0,03$.

c. [3] Se fossero ignote le varianze, sarebbe possibile fornire una funzione test opportuna per il sistema di ipotesi dato? Occorrono assunzioni ulteriori?

d. [3] Si immagini ora che i valori riportati in tabella si riferiscano alle varianze campionarie per i due campioni. Si rifiuterebbe al livello 0.05 l'ipotesi di uguaglianza delle due varianze?

2. La v.a. X, tempo di attesa tra due arrivi di un cliente a uno sportello di un ufficio espresso in minuti, è distribuita secondo un'esponenziale di parametro λ.

a. [3] E' noto che la probabilità che il tempo tra due arrivi sia maggiore di 5 minuti è 0.75; si dica quanto vale λ.

b. [3] Si determini la mediana di X.

c. [3] Qual è la probabilità che occorrano più di 30 ore per osservare 100 arrivi.

3. La variabile X è distribuita come una $N(\mu, 100)$. Si estraggono casualmente 200 valori x_i dalla popolazione ma si registra solo se ciascun valore estratto è superiore a 230. Per chiarezza si definisca la variabile Y tale che $y_i = 1$ se $x_i > 230$. Per il campione dato $\sum_i^n y_i = 120$.

a. [3] Si fornisca uno stimatore di massima verosimiglianza della probabilità che $X > 230$.

b. [3] Si ottenga la stima (di massima verosimiglianza) per μ.

c. [3] Dare una valutazione approssimata della varianza dello stimatore ottenuto al punto precedente.

4. Si dispone di un campione casuale di 300 clienti presso una catena di supermercati e per ciascuno di essi si misura la spesa effettuata in euro X. Nella tabella che segue sono riportati le frequenze osservate per classi di valori di $Y = logX$. Si supponga che la varianza di Y sia nota e pari a 2.

log(spesa)	meno di 3	3–4	4–5	5–7	oltre 7
frequenza	45	62	94	71	38

a. [3] Il direttore della catena ha fornito il valore totale della spesa per i 300 clienti $\sum_{i=1}^{300} x_i$ che è stata pari a 36000 euro. Lo statistico A non conoscendo la media della variabile Y decide di stimarla come il logaritmo di 36000/300 evitando di usare l'approssimazione dovuta la raggruppamento in classi. Lo statistico B dice invece che è meglio stimarla direttamente dalla tabella usando i valori centrali delle classi come valori rappresentativi per tutti i clienti della classe. Chi dei due ha ragione e perchè?

b. [6] Verificare l'ipotesi che la variabile Y sia distribuita come una gaussiana al livello $\alpha = 0.01$.