Politecnico di Milano - Scuola di Ingegneria Industriale

II Appello di Statistica per Ingegneria Energetica
5 settembre 2011

Cognome, Nome e Numero di matricola:

Problema 1. Nella cittadina di Los Eisley è attivo da meno di due anni il corso di studio in Ingegneria dello Svago. È il momento di trarre i primi bilanci di questa attività e pertanto il rettore Hutt si è fatto consegnare i voti di laurea di tutti i 100 studenti che hanno terminato la laurea triennale. Questi voti sono riportati in tabella, dove il 110 e lode è stato codificato con 111:

Voto	**Freq. Assoluta**
66	1
70	1
80	5
85	3
90	8
94	16
97	12
100	10
105	12
108	10
110	14
111	8

(a) Si calcolino minimo, massimo, mediana, primo e terzo quartile dei voti di laurea.

(b) Si disegni il boxplot. Quali commenti si possono fare sulla distribuzione dei voti?

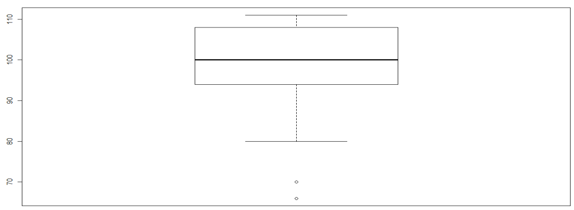
(c) Si individuino eventuali outlier. Se venissero eliminati, cambierebbe il valore della mediana?

(d) A quali percentili corrisponde il voto 110?

Risultati:

(a)
Minimo= 66, Massimo= 111, Mediana= 100, Primo quartile= 94, Terzo quartile= 108.

(b)



La distribuzione presenta una leggera asimmetria a sinistra, evidenziata dal baffo inferiore più lungo e dalla presenza di due outlier inferiori.

(c)
I due outlier inferiori sono i voti 66 e 70. La mediana non cambierebbe anche eliminando le due osservazioni corrispondenti.

(d)
Il voto 110 corrisponde ai percentili dal 78% al 92%, estremi esclusi. Infatti

$$x_{(78)} = 108, \qquad x_{(79)} = \ldots = x_{(92)} = 110, \qquad x_{(93)} = 111,$$

per cui

$$q_\alpha = 110 \quad \forall \alpha \in (0.78\,,\, 0.92),$$

mentre

$$q_{0.78} < 110, \qquad q_{0.92} > 110.$$

Problema 2. Volete misurare una resistenza R ed avete a disposizione uno strumento digitale di risoluzione 0.1 ohm. La misura di R vi dà una lettura di 2 ohm.

(a) Quale distribuzione è ragionevole supporre per i possibili valori di R?

(b) Qual è il valore atteso di R?

(c) Quanto vale la sua deviazione standard σ_R?

(d) Calcolare $P\Big(2 - \sigma_R < R < 2 + \sigma_R\Big)$.

Adesso dovrete collegare 33 resistenze R_k, $k = 1, \ldots, 33$, che danno tutte la stessa lettura di 2 ohm ed i cui loro valori effettivi possono essere ritenuti indipendenti fra di loro. Per prima cosa dovete collegare tutte le resistenze in serie, ottenendo una resistenza totale $S = \sum_{k=1}^{33} R_k$.

(e) Qual è il valore atteso di S?

(f) Quanto vale la sua deviazione standard σ_S?

(g) Quale distribuzione, eventualmente approssimata, possiamo supporre per i possibili valori di S?

(h) Calcolare, eventualmente in modo approssimato, $P\Big(66 - \sigma_S < S < 66 + \sigma_S\Big)$.

Infine dovete collegare in parallelo solo R_1 ed R_2, ottenendo una resistenza totale $T = \dfrac{R_1\, R_2}{R_1 + R_2}$. Utilizzando il metodo delta si valutino:

(i) il valore atteso di T,

(j) la sua deviazione standard σ_T.

Risultati.

(a) $R \sim U(1.95\,, 2.05)$

(b) $\mu_R = 2$

(c) $\sigma_R = \sqrt{\dfrac{0.1^2}{12}} = \dfrac{0.1}{2\sqrt{3}} = 0.028867513$

(d) $P\Big(2 - \sigma_R < R < 2 + \sigma_R\Big) = \displaystyle\int_{2-\sigma_R}^{2+\sigma_R} \frac{1}{0.1}\, \mathrm{d}r = \frac{2\sigma_R}{0.1} = \frac{1}{\sqrt{3}} = 0.57735$

(e) $\mu_S = \sum_{k=1}^{33} \mu_{R_k} = 33 \cdot 2 = 66$

(f) $\sigma_S = \sqrt{\operatorname{Var}\left(\displaystyle\sum_{k=1}^{33} R_k\right)} = \sqrt{\displaystyle\sum_{k=1}^{33} \operatorname{Var}(R_k)} = \sqrt{33}\,\sigma_R = 0.05\,\sqrt{11} = 0.16583124$ per l'indip. delle R_k.

(g) $S = \sum_{k=1}^{33} R_k \sim N(66\,,\, 0.0275)$ per il TCL, supponendo il campione sufficientemente numeroso.

(h) $P\Big(66 - \sigma_S < S < 66 + \sigma_S\Big) = P\Big(\dfrac{|S - 66|}{\sigma_S} < 1\Big) \simeq 2\Phi(1) - 1 = 0.6827$

(i) $\mu_T \simeq \dfrac{\mu_{R_1}\, \mu_{R_2}}{\mu_{R_1} + \mu_{R_2}} = 1$

(j) Scrivendo $T = \dfrac{R_1\, R_2}{R_1 + R_2} = h(R_1, R_2)$, otteniamo

$$\begin{aligned}\sigma_T = \sqrt{\operatorname{Var} T} &\simeq \sqrt{\Big(\partial_1 h(\mu_{R_1}, \mu_{R_2})\Big)^2 \operatorname{Var} R_1 + \Big(\partial_2 h(\mu_{R_1}, \mu_{R_2})\Big)^2 \operatorname{Var} R_2} \\ &= \sqrt{\Big(\frac{\mu_{R_2}}{\mu_{R_1} + \mu_{R_2}}\Big)^4 \operatorname{Var} R_1 + \Big(\frac{\mu_{R_2}}{\mu_{R_1} + \mu_{R_2}}\Big)^4 \operatorname{Var} R_2} = \sqrt{\frac{1}{2^3}}\,\sigma_R = 0.010206207\end{aligned}$$

Problema 3. L'azienda PharmaENG sta per mettere in commercio un farmaco antistaminico di nuova generazione. Si sospetta però che il farmaco possa avere l'indesiderato effetto collaterale di un aumento della pressione sistolica. Viene quindi somministrato il farmaco a 12 volontari registrando i valori di pressioni sanguigne sistoliche immediatamente prima e dopo la somministrazione, riportati nella seguente tabella.

Volontario	1	2	3	4	5	6	7	8	9	10	11	12
Prima	110	128	132	103	143	125	121	114	129	141	119	128
Dopo	118	123	143	114	145	133	118	124	138	148	121	129

(a) Stabilire mediante un opportuno test statistico se i dati portano a concludere che il farmaco produca un innalzamento della pressione (media) sistolica, specificando le opportune ipotesi da assumere sulla distribuzione delle variabili in esame.

(b) Calcolare un intervallo di confidenza bilaterale di livello 95% per la differenza delle medie della pressione sistolica prima e dopo la somministrazione dell'antistaminico.

Risultati.

(a) Si deve utilizzare un t-test unilatero per dati accoppiati assumendo che la differenza tra la pressione prima e dopo sia distribuita secondo una legge gaussiana. P-value $= 0.0042$. Si può ritenere che il farmaco provochi un innalzamento della pressione media sistolica.

(b) IC $= [-8.578742; -1.587925]$

Problema 4. Un produttore di batterie produce pile di diverse qualità A, B, C, D, E, F (in ordine decrescente di durata) con percentuali rispettivamente del 25%, 30%, 25%, 10%, 7% e 3%. Il produttore afferma di distribuire le pile ai suoi clienti nelle stesse proporzioni di produzione. Uno dei clienti ha però l'impressione di ricevere troppi pezzi di scarsa qualità. Decide pertanto di controllare il livello qualitativo delle prossime n pile e vi chiede aiuto per verificare sulla base dei risultati campionari la politica di distribuzione del produttore.

1. Impostare un opportuno test di ipotesi per affrontare il problema posto dal cliente. Esplicitate la popolazione oggetto di indagine statistica, le ipotesi nulla e alternativa, la regione critica di livello α, le condizioni necessarie all'utilizzo del test.

Il cliente riceve 100 pile.

2. Sono verificate le condizioni necessarie all'utilizzo del test introdotto in (a)? In caso contrario, non essendo possibile aumentare la numerosità del campione, modificate opportunamente il test.

Il controllo delle 100 pile dà i seguenti risultati:

Qualità batterie	A	B	C	D	E	F
# batterie	22	29	27	12	6	4

3. Calcolare il p-value dei dati.

4. Se il cliente vuole evitare di accusare ingiustamente il produttore, limitando al 10% la probabilità di un tale sconveniente errore, cosa può concludere circa la politica di distribuzione del produttore?

Risultati.

1. La variabile aleatoria di interesse è X = qualità di una pila ricevuta dal cliente. Questa assume i valori $i = A, B, C, D, E, F$ con distribuzione $p_i = P(X = i)$ che dobbiamo ritenere incognita.

   Per verificare se c'è evidenza sperimentale del fatto che il produttore non rispetti le percentuali di distribuzione dichiarate p_i^0 vogliamo eseguire un test per

   $$H_0 : p_i = p_i^0 \quad \forall\ i = A, \dots, F, \qquad \text{contro} \qquad H_1 : p_i \neq p_i^0 \quad \text{per qualche } i = A, \dots, F,$$

   dove
   $$p_A^0 = 0.25, \quad p_B^0 = 0.3, \quad p_C^0 = 0.25, \quad p_D^0 = 0.1, \quad p_E^0 = 0.07, \quad p_F^0 = 0.03.$$

   Per rifiutare sulla base di un campione casuale $X_1, \dots, X_n$ introduciamo un test χ^2 di adattamento di livello α usando per classi i sei possibili valori di X, ovvero la regione critica

   $$Q_n = \sum_{i=A}^{F} \frac{(N_i - np_i^0)^2}{np_i^0} > \chi^2_\alpha(5)$$

   dove N_i è la frequenza assoluta del caso i e $\chi^2_\alpha(5)$ è il punto percentuale di ordine α di una $\chi^2(5)$.

   È necessario che il campione sia casuale e numeroso: il livello nominale del test è una buona approssimazione di quello reale solo se $np_i^0 > 5$ per ogni i.

2. Per $n = 100$ le frequenze attese delle classi sono

   $$np_A^0 = 25, \quad np_B^0 = 30, \quad np_C^0 = 25, \quad np_D^0 = 10, \quad np_E^0 = 7, \quad np_F^0 = 3.$$

   Dato che $np_F^0 = 3 < 5$ si possono unire la classe E e la classe F ottenendo

Classi i	{A}	{B}	{C}	{D}	{E,F}
$100\,p_i^0$	25	30	25	10	10

   La corrispondente regione critica è

   $$Q_n = \sum_{i=1}^{5} \frac{(N_i - np_i^0)^2}{np_i^0} > \chi^2_\alpha(4)$$

   dove ora N_i e p_i^0 sono, rispettivamente, frequenza assoluta e probabilità dichiarata della classe i.

3. Con i nostri dati otteniamo

Classi i	{A}	{B}	{C}	{D}	{E,F}
N_i	22	29	27	12	10
$100\,p_i^0$	25	30	25	10	10

   da cui $Q_{100} = 0.953$. Dalle tavole si ricava che il p-value dei dati è in $(0.9\,,\,0.95)$.

4. Per evitare di accusare ingiustamente il produttore, limitando al 10% la probabilità di un tale sconveniente errore, si deve eseguire il test a livello $\alpha = 0.1$. Essendo tale livello inferiore al p-value, non possiamo rifiutare H_0 al livello del 10%. Pertanto il cliente non può affermare che il produttore usi una politica di distribuzione diversa da quella dichiarata.