Politecnico di Milano - Scuola di Ingegneria Industriale e dell'Informazione

III APPELLO DI STATISTICA PER INGEGNERIA ENERGETICA
23 luglio 2015

Cognome, Nome e Numero di matricola:

Problema 1. In Italia viene organizzata una scuola estiva di ingegneria elettrica alla quale partecipano 180 studenti conseguendo le seguenti votazioni finali:

Voto	Frequenza Assoluta
24	4
26	26
27	45
28	84
29	3
30	10
30L	8
	180

Nel seguito si consideri il 30L come un 31.

(a) Completare la tabella di distribuzione di frequenza con la Frequenza Assoluta Cumulata.

(b) Rappresentare la distribuzione dei dati con un istogramma.

(c) Rappresentare la distribuzione dei dati con un boxplot.

(d) Per quali α il voto 30L risulta punto percentuale di ordine α?

Era previsto che gli studenti top 3% fossero ammessi al programma speciale AV, ma l'eccessiva presenza di 30L rende impossibile individuare tali studenti. Gli organizzatori decidono quindi di aggiungere un super esame finale per stabilire la graduatoria definitiva degli studenti. Se sarà possibile individuare gli studenti top 3%, il programma AV partirà, altrimenti sarà cancellato. Il giovane Anthony vuole partecipare al programma AV, è fra gli studenti che hanno preso 30L, non dubita di riprendere 30L, non dubita che anche Bruce riprenderà 30L, ritiene che gli altri sei 30L siano invece frutto dei suoi suggerimenti, confida che il solitario Bruce rimarrà come sempre concentrato sul proprio esame senza suggerire a nessuno.

(e) A quanti compagni Anthony potrà dare suggerimenti senza timore di cancellare il programma AV?

Risultati.

(a)

Voto	Frequenza Assoluta	Frequenza Assoluta Cumulata
24	4	4
26	26	30
27	45	75
28	84	159
29	3	162
30	10	172
30L	8	180
	180	

(b) Dato che $1 + \log_2 180 = 8.5$ mentre $\sqrt{180} = 13.4$, teniamo come classi i singoli voti, aggiungendo ovviamente il 25, arrivando così a 8 classi. Si trova quindi il seguente istogramma

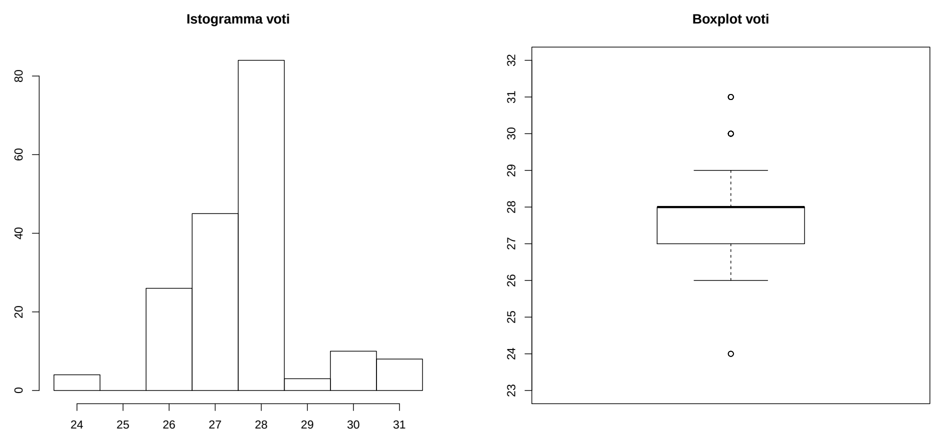

Figura 1: Istogramma e boxplot come richiesto nei punti (b) e (c)

(c) Q_1: $180/4 = 45 \implies Q_1 = \dfrac{x_{(45)} + x_{(46)}}{2} = 27$ Q_2: $180/2 = 90 \implies Q_1 = \dfrac{x_{(90)} + x_{(91)}}{2} = 28$

Q_3: $180 \cdot 3/4 = 135 \implies Q_1 = \dfrac{x_{(135)} + x_{(136)}}{2} = 28$

$\text{IQR} = Q_3 - Q_1 = 1$

Voti erratici inferiori: 24. Voti erratici superiori: 30, 30L.

Si trova quindi il boxplot sopra riportato.

(d) $p_\alpha = q_{1-\alpha} = 30\text{L} \iff 180 \cdot (1 - \alpha) \geq 172 \iff \alpha \leq \dfrac{8}{180} = 0.0444 = 4.44\%$.

(e) $180 \cdot 3\% = 5.4$ per cui sarà possibile individuare gli studenti top 3% solo se i 30L saranno 5 al massimo.

Quindi Anthony potrà dare suggerimenti a massimo 3 compagni.

Problema 2. Un filo di ragnatela di Peter ha lunghezza casuale X con distribuzione, in **decametri**,

$$f(x) = \frac{\alpha}{x^4} I_{[\beta,+\infty)}(x),$$

dove i parametri α e β sono da determinare.

(a) Quali condizioni devono soddisfare α e β affinché f risulti una densità di probabilità?

(b) Determinare α e β sapendo che i fili di ragnatela di Peter sono mediamente lunghi 1.5.
Tracciare quindi un grafico qualitativo della densità f.

(c) Calcolare la varianza della lunghezza dei fili di ragnatela di Peter.

Unendo più fili, le cui lunghezze sono indipendenti fra di loro, volete coprire una lunghezza pari a 1 Km.

(d) Quale distribuzione ha la lunghezza Y di n fili di ragnatela di Peter?

(e) Quanti fili vi servono per coprire 1 Km con una probabilità superiore al 75%?

Risultati.

(a)

$$\begin{cases} f(r) \geq 0 \\ \int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1 \end{cases} \iff \begin{cases} \alpha \geq 0 \\ 1 = \alpha \int_{\beta}^{+\infty} \frac{1}{x^4}\,\mathrm{d}x = \dfrac{\alpha}{3\beta^3} \end{cases} \iff \begin{cases} \beta > 0 \\ \alpha = 3\beta^3 \end{cases}$$

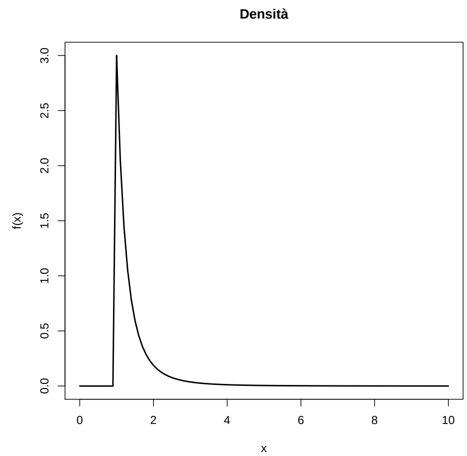

Figura 2: Grafico della densità

(b)

$$\frac{3}{2} = \mathbb{E}[X] = 3\beta^3 \int_{\beta}^{+\infty} x\,\frac{1}{x^4}\,\mathrm{d}x = \frac{3}{2}\,\beta \iff \beta = 1 \implies \alpha = 3$$

(c)

$$\mathbb{E}[X^2] = 3 \int_{1}^{+\infty} x^2\,\frac{1}{x^4}\,\mathrm{d}x = 3 \implies \mathrm{Var}(X) = \frac{3}{4} = 0.75$$

(d) $Y = \sum_{k=1}^{n} X_k \simeq N\left(\frac{3}{2}\,n\ ,\ \frac{3}{4}\,n\right)$ se n abbastanza grande ($n \geq 40$ per esempio).

(e)

$$0.75 < \mathbb{P}(Y \geq 100) \simeq 1 - \Phi\left(\frac{100 - \frac{3}{2}\,n}{\sqrt{\frac{3}{4}\,n}}\right)$$

$$\frac{100 - \frac{3}{2}\,n}{\sqrt{\frac{3}{4}\,n}} < -z_{0.25} = -0.674$$

$$\sqrt{n} > 8.36 \iff n \geq 70 \text{ (coerente con l'ipotesi } n \geq 40)$$

Problema 3. Bruce e Clark si stanno allenando al meglio per prepararsi alla sfida che li vedrà protagonisti il prossimo anno. Tra i numerosi esercizi fisici a cui si sottopongono, uno consiste nel cercare di migliorare la potenza del proprio pugno sferrato con il braccio dominante. Alfred, contrariamente a Bruce, è molto preoccupato e vuole essere sicuro che il suo pupillo arrivi allo scontro senza troppe preoccupazioni. A tal proposito, raccoglie 63 prove (n_B) di forza di Bruce in modo da confrontarle con le 75 (n_C) che è riuscito a recuperare, sottraendole di soppiatto, dalla palestra dove si allena Clark. Si riportano media e varianza campionaria per i due campioni in opportune unità di misura: $\overline{x}_B = 100.6759$ e $s_B^2 = 15.4058$, $\overline{x}_C = 99.6255$ e $s_C^2 = 17.5434$. I test di Shapiro - Wilk eseguiti su i due campioni danno come p-value i seguenti valori: $p_B = 0.4439$ e $p_C = 0.7713$. Di seguito si riportano il boxplot dei dati unitamente ai Normal Probability Plot.

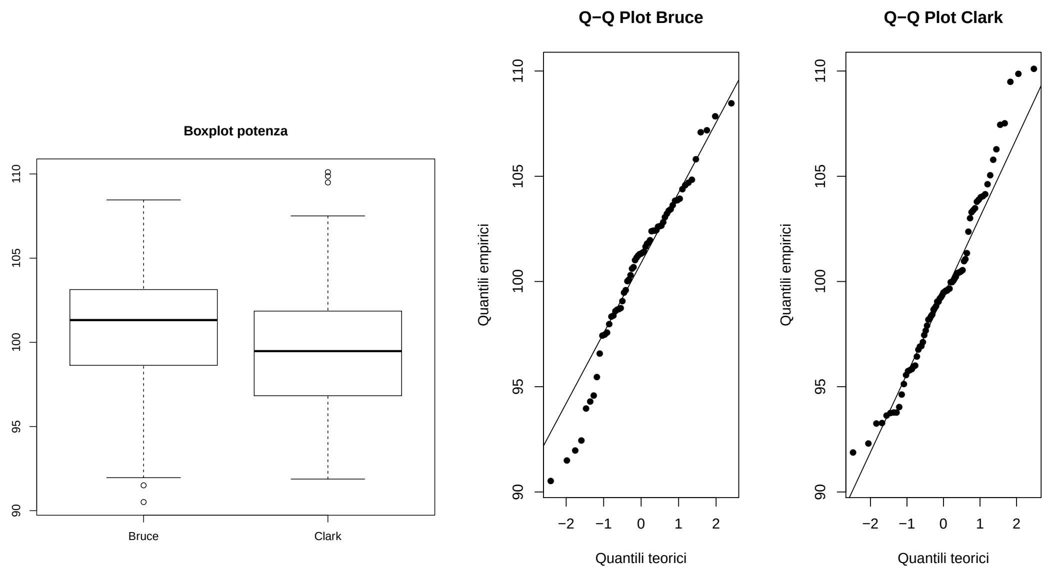

Figura 3: Boxplot e Normal Probability Plot della potenza del pugno di Bruce e di Clark

(a) Si esegua un test ad un livello del 10% per stabilire se la varianza della potenza dei pugni di Bruce possa essere ritenuta uguale alla varianza della potenza dei pugni di Clark. Si esplicitino: ipotesi nulla, ipotesi alternativa, regione critica, esito del test. Esplicitare e discutere le condizioni di applicabilità del test usato.

(b) In caso di risposta affermativa, si stimi puntualmente la varianza comune delle due potenze.

(c) Si imposti un test per provare statisticamente che la potenza media dei pugni di Bruce è superiore alla potenza media dei pugni di Clark. Si esplicitino: ipotesi nulla, ipotesi alternativa, errore di primo tipo, regione critica di livello α. Esplicitare e discutere le condizioni di applicabilità del test introdotto.

(d) Calcolare il p-value dei dati per il test appena introdotto.

(e) Siete d'accordo con Alfred e ritenete che Bruce debba allenarsi di più? Giustificare la risposta.

Risultati. Siano

$$X_B = \text{ la potenza di un pugno di Bruce,} \qquad \mu_B = \mathbb{E}[X_B], \qquad \sigma_B^2 = \mathrm{Var}(X_B),$$

$$X_C = \text{ la potenza di un pugno di Clark,} \qquad \mu_C = \mathbb{E}[X_C], \qquad \sigma_C^2 = \mathrm{Var}(X_C),$$

(a)

$$H_0 : \sigma_B^2 = \sigma_C^2 \qquad \text{vs} \qquad H_1 : \sigma_B^2 \neq \sigma_C^2$$

$$R_{0.1} : \left\{ f_0 = \frac{s_B^2}{s_C^2} < f_{0.95;62,74} = \frac{1}{f_{0.05;74,62}} = \frac{1}{1.5031} = 0.6653 \right\} \cup \left\{ f_0 = \frac{s_B^2}{s_C^2} > f_{0.05;62,74} = 1.4906 \right\}$$

Poiché

$$f_0 = \frac{s_B^2}{s_C^2} = \frac{15.4058}{17.5434} = 0.8782,$$

entrambe le disuglianze non sono soddisfatte, quindi accetto H_0: varianze uguali.

Approssimando i gradi di libertà della Fisher, si approssima la regione critica

$$R_{0.1} : \left\{ f_0 = \frac{s_B^2}{s_C^2} < f_{0.95;60,60} = \frac{1}{f_{0.05;60,60}} = \frac{1}{1.53} = 0.6536 \right\} \cup \left\{ f_0 = \frac{s_B^2}{s_C^2} > f_{0.05;60,60} = 1.53 \right\}$$

trovando stessa conclusione.

Il test può essere eseguito, tenendone sotto controllo il livello, se X_B e X_C sono normali. La validità di questa ipotesi è confermata dai dati campionari, sia alla luce dei NPP, sia alla luce degli alti p-value dei test di Shapiro-Wilk.

(b) Le due varianze sono risultate uguali quindi lo stimatore congiunto di $\sigma^2 = \sigma_B^2 = \sigma_C^2$ è dato da:

$$s_p^2 = \frac{s_B^2(n_B - 1) + s_C^2(n_C - 1)}{n_B + n_C - 2} = 16.5689.$$

(c)

$$H_0 : \mu_B \leq \mu_C \qquad \text{vs} \qquad H_1 : \mu_B > \mu_C$$

Errore di primo tipo: Bruce non aumenta gli allenamenti nonostante $\mu_C \geq \mu_B$.

$$R_\alpha : t_0 = \frac{\overline{x}_B - \overline{x}_C}{s_p\sqrt{\frac{1}{n_B} + \frac{1}{n_C}}} > t_{\alpha;n_B+n_C-2}$$

Il test può essere eseguito, tenendone sotto controllo il livello, grazie alla normalità di X_B e X_C (vedi sopra).

(d) Per i dati raccolti

$$t_0 = 1.50996 = t_{\alpha;136} \simeq z_\alpha \implies \alpha \simeq 1 - \Phi(1.51) = 0.07$$

Il p-value risulta quindi tra 0.05 e 0.1. Il rifiuto di H_0 dipende pertanto dalla scelta del livello di confidenza con il quale si vuole effettuare il test, ovvero da quanto si vuole rischiare un errore di primo tipo.

Agli usuali livelli del 5% e dell'1% non c'è ragione statistica per rifiutare H_0 (Bruce deve allenarsi di più), mentre con un livello del 10% H_0 sarà rifiutata (Bruce può stare sereno).