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

V Appello di Statistica per Ingegneria Energetica
21 settembre 2015

Cognome, Nome e Numero di matricola:

Problema 1. Si consideri un numero casuale X di legge normale $N(\mu, \sigma^2)$ con quartili

$$q_{0.25} = Q_1 = 9, \qquad q_{0.75} = Q_3 = 11.$$

Calcolare:

(a) la media μ,

(b) la probabilità $\mathbb{P}\Big(|X - \mu| > 1\Big)$,

(c) la varianza σ^2,

(d) il momento secondo $\mathbb{E}[X^2]$,

(e) la probabilità $\mathbb{P}\Big(|X - \mu| > 2\Big)$.

Risultati. Sia come al solito $Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$

(a) $\mu = 10$ per la simmetria delle densità gaussiane,

(b) $\mathbb{P}\Big(|X-\mu| > 1\Big) = \mathbb{P}(X \leq q_{0.25}) + \mathbb{P}(X \geq q_{0.75}) = 0.5$,

(c) $\sigma^2 = 2.2013$. Infatti:

$$0.25 = \mathbb{P}(X > 11) = \mathbb{P}\left(Z > \frac{1}{\sigma}\right) \quad\Longleftrightarrow\quad \frac{1}{\sigma} = z_{0.25} = 0.674 \quad\Longleftrightarrow\quad \sigma^2 = (1.4837)^2 = 2.2013$$

(d) $\mathbb{E}[X^2] = \mu^2 + \sigma^2 = 102.2013$,

(e) $\mathbb{P}\Big(|X-\mu| > 2\Big) = \mathbb{P}\Big(|Z| > \dfrac{2}{\sigma}\Big) = \mathbb{P}(|Z| > 1.348) \simeq 2\Big(1 - \Phi(1.35)\Big) = 0.177016$.

Problema 2. I cristalli di dilitio costituiscono la più importante fonte energetica per il corretto funzionamento dei motori a curvatura presenti sulla maggior parte delle navi stellari. Il minerale è estremamente raro e inoltre la sua purezza X deve essere piuttosto alta ($X > 10$) per poter essere impiegato efficientemente. Il capo ingegnere O'Brian sta valutando se iniziare o meno le operazioni di estrazione da un giacimento appena scoperto. Ordina quindi l'estrazione di un campione casuale di 50 cristalli di dilitio per poterne misurare la purezza. Di seguito si riportano media campionaria, varianza campionaria, tabella di distribuzione frequenza, istogramma e NPP delle 50 purezze osservate.

$$\overline{x}_{50} = 11.0791, \qquad s_{50}^2 = 1.0289.$$

Classe	Frequenza Assoluta
$[9, 10]$	7
$(10, 11]$	19
$(11, 12]$	16
$(12, 13]$	6
$(13, 14]$	2

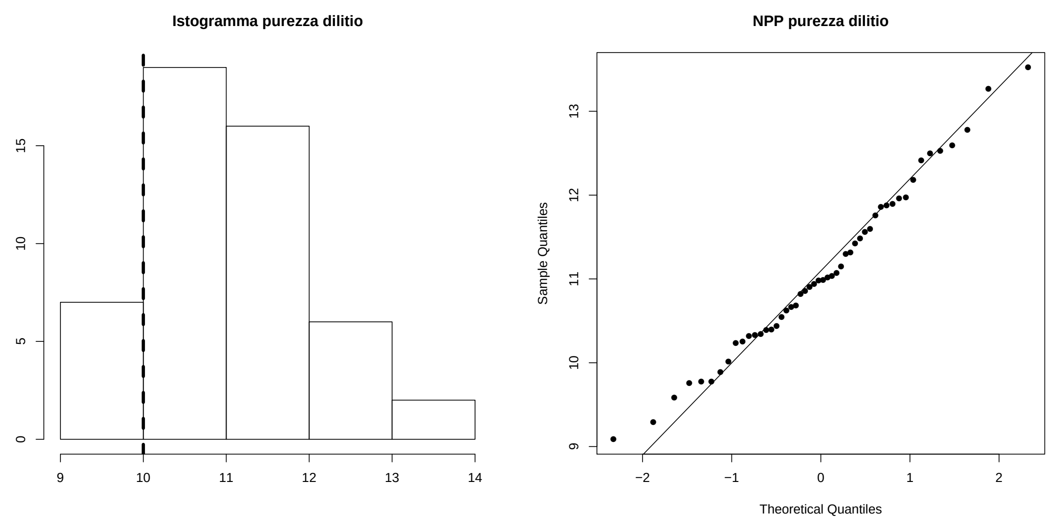

Figura 1: Istogramma e NPP della purezza di 50 cristalli di dilitio provenienti dal nuovo giacimento

Si risponda ai seguenti quesiti, adoperando gli opportuni strumenti statistici, e discutendone ogni volta l'utilizzabilità.

1. Il capo O'Brian può affermare che tutti i cristalli del giacimento hanno purezza > 10?
2. Il capo O'Brian può affermare che almeno l'85% dei cristalli del giacimento ha purezza > 10?
3. Il capo O'Brian può affermare che i cristalli del giacimento hanno purezza media > 10?

Risultati.

1. No, ovviamente: il giacimento contiene cristalli di purezza $X \leq 10$, come mostrato dal campione.

2. Si deve eseguire un test sulla proporzione p di cristalli di purezza > 10 in tutto il giacimento.

   Le ipotesi statistiche da verificare sono:

$$H_0 : \; p < 0.85 \qquad \text{vs.} \qquad H_1 : \; p \geq 0.85$$

   Grazie alla numerosità del campione raccolto (che contiene più di 5 successi e più di 5 insuccessi) possiamo usare la regione critica di livello α

$$R_\alpha : \hat{p} > 0.85 + \sqrt{\frac{0.85(1-0.85)}{50}}\, z_\alpha$$

   dove $\hat{p}$ è la stima puntuale di p fornita dal campione, che vale $\hat{p} = \frac{19+16+6+2}{50} = 43/50 = 0.86$.

   Calcoliamo quindi il p-value dei dati raccolti:

$$\hat{p} = 0.85 + \sqrt{\frac{0.85(1-0.85)}{50}}\, z_\alpha$$

$$\text{p-value} = \alpha = 1 - \Phi\left(\frac{0.86-0.85}{\sqrt{\frac{0.85(1-0.85)}{50}}}\right) = 1 - \Phi(0.1980295) = 1 - 0.5793 = 0.4207$$

   Agli usuali livelli di significatività non è quindi possibile rifiutare H_0: nonostante il campione contenga l'86% di cristalli di purezza > 10, il capo O'Brian non può affermare che almeno l'85% dei cristalli del giacimento abbia purezza > 10 (conclusione debole).

3. Si deve eseguire un test sulla purezza media μ dei cristalli di tutto il giacimento.

   Le ipotesi statistiche da verificare sono:

$$H_0 : \; \mu \leq 10 \qquad \text{vs.} \qquad H_1 : \; \mu > 10$$

   Grazie alla gaussianità della purezza dei cristalli del giacimento (come confermato dal NPP del campione) e/o alla numerosità del campione raccolto (50 osservazioni) possiamo usare la regione critica di livello α per popolazioni normali a varianza incognita

$$R_\alpha : \overline{x}_{50} > 10 + \frac{s_{50}}{\sqrt{50}}\, t_\alpha(49)$$

   Calcoliamo quindi il p-value dei dati raccolti:

$$\overline{x}_{50} = 10 + \frac{s_{50}}{\sqrt{50}}\, t_\alpha(49) \qquad \Longrightarrow \qquad t_\alpha(49) = 7.5224$$

   Usando le tavole si trova

$$\text{p-value} = \alpha << 0.0005,$$

   mentre usando un calcolatore si trova

$$\text{p-value} = \alpha = 5.167 \cdot 10^{-10}.$$

   Agli usuali livelli di significatività (e non solo!) è quindi possibile rifiutare H_0: il capo O'Brian può affermare che i cristalli del giacimento hanno purezza media > 10 (conclusione forte).

Problema 3. Nei Laboratori di Fisica Sperimentale di Viale Gran Sasso alcune matricole di Ingegneria Matematica vogliono validare statisticamente la legge oraria di caduta di un grave. Posizionano pertanto un oggetto di massa m ad una certa altezza dal suolo y_0 e lo lasciano cadere sotto l'azione del suo solo peso. Detta y l'altezza dal suolo del grave al tempo t, teoricamente dovrebbe quindi valere

$$y = y_0 - \frac{1}{2} g\, t^2.$$

Gli studenti usano un grave di massa $m = 1\,\mathrm{kg}$ e agli istanti (in secondi) $t_i = 0.1, 0.2, \ldots, 1$ misurano l'altezza dal suolo del grave (in metri). Tuttavia i risultati y_i delle loro 10 misure sono affetti da errore casuale gaussiano di media 0.

(a) Introdurre un modello lineare gaussiano che leghi y, t e l'errore di misura ϵ.

Con le loro misure gli studenti stimano puntualmente l'accelerazione di gravità con $\widehat{g} = 10\ \mathrm{m/s^2}$ e l'altezza iniziale con $\widehat{y_0} = 5$ m.

(b) Fornire una stima puntuale dei coefficienti del modello lineare.

(c) Sapendo che la somma degli scarti dalla media delle altezze dal suolo è $SS_{yy} = 27.659\ \mathrm{m}^2$, calcolare l'indice R^2 del modello lineare.

(d) Fornire una stima puntuale della varianza dell'errore di misura.

(e) Fornire un intervallo di previsione al $\alpha = 5\%$ per l'altezza dal suolo a $t = 0.65$ s.

Risultati.

(a) $Y = y_0 - \dfrac{1}{2} g\ t^2 + \epsilon = \beta_0 + \beta_1\ x + \epsilon, \qquad x = t^2, \qquad \epsilon \sim N(0, \sigma^2).$

(b) $\widehat{\beta}_1 = -\dfrac{1}{2}\widehat{g} = -5, \qquad \widehat{\beta}_0 = \widehat{y_0} = 5.$

(c) $R^2 = \widehat{\beta}_1^2 \dfrac{SS_{t^2t^2}}{SS_{yy}}$. Dal calcolo diretto: $SS_{t^2t^2} = 1.05105$. Dunque $R^2 = 25\ \dfrac{1.05105}{27.659} = 0.95.$

(d) $\widehat{\sigma}^2 = \dfrac{SS_{res}}{8} = \dfrac{(1-R^2)SS_{yy}}{8} = 0.1729\ \mathrm{m}^2.$

(e) $IP(y) = \widehat{y} \pm t_{1-\alpha/2}(8) \sqrt{\widehat{\sigma}^2 \left(1 + \dfrac{1}{n} + \dfrac{\left(t^2 - \overline{t^2}\right)^2}{SS_{t^2t^2}}\right)} = 2.8875 \pm 1.006.$