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

Appello di Statistica per Ingegneria Energetica
22 settembre 2014

Cognome, Nome e Numero di matricola:

Problema 1. Sia X il numero aleatorio di regali che Leonardo riceverà per il suo primo compleanno. Consideriamo le funzioni F e G, dipendenti dai parametri $\alpha, \beta \in \mathbb{R}$:

$$F(x) = \begin{cases} 0, & x < 1 \\ \frac{1}{6}, & 1 \leq x < 2 \\ \alpha, & 2 \leq x < 3 \\ \frac{5}{6}, & 3 \leq x < 4 \\ 1, & x \geq 4 \end{cases}, \qquad G(x) = \begin{cases} 0, & x < 1 \\ \frac{1}{3}, & 1 \leq x < 2 \\ \beta, & 2 \leq x < 4 \\ \frac{1}{2}, & x \geq 4 \end{cases},$$

a) Per quali valori di $\alpha, \beta \in \mathbb{R}$ le funzioni F e G possono essere funzioni di ripartizione di X?

Sia ora F la funzione di ripartizione di X.

b) Scrivere la densità di probabilità discreta (funzione di massa di probabilità) di X.

c) Sapendo che la probabilità che Leonardo riceva almeno 3 regali per il suo compleanno è pari a $\frac{2}{3}$, calcolare quanto vale il parametro α.

d) Calcolare il valor atteso e la varianza di X.

e) Qual è la probabilità che Leonardo, tra 40 anni, abbia ricevuto meno di 100 regali di compleanno? Si considerino le variabili $X_1, \ldots, X_{40}$ che contano i regali ricevuti nei vari anni indipendenti e con la stessa distribuzione di X.

Risultati.

a) La funzione F è funzione di ripartizione per $\frac{1}{6} \leq \alpha \leq \frac{5}{6}$ (infatti è non decrescente, continua da destra, con $\lim_{x\to-\infty} F(x) = 0$ e $\lim_{x\to+\infty} F(x) = 1$).

La funzione G non può essere una funzione di ripartizione per nessun valore di β in quanto $\lim_{x\to+\infty} G(x) = 1/2 \neq 1$.

b) $p(1) = \dfrac{1}{6}, \qquad p(2) = \alpha - \dfrac{1}{6}, \qquad p(3) = \dfrac{5}{6} - \alpha, \qquad p(4) = \dfrac{1}{6}.$

c) $\dfrac{2}{3} = P(X \geq 3) = p(3) + p(4) = \dfrac{5}{6} - \alpha + \dfrac{1}{6} = 1 - \alpha \implies \alpha = \dfrac{1}{3}.$

d)

$$E[X] = \sum_{x=1}^{4} x p(x) = \frac{8}{3},$$

$$Var(X) = E[X^2] - (E[X])^2 = \sum_{x=1}^{4} x^2 p(x) - \left(\frac{8}{3}\right)^2 = 8 - \frac{64}{9} = \frac{8}{9}.$$

e) Consideriamo la variabile aleatoria $Y = \sum_{i=1}^{40} X_i$ che conta il numero di regali di compleanno ricevuti in 40 anni. Siccome $X_1, \ldots, X_{40}$ sono iid distribuite come X e $n = 40$ è grande, possiamo utilizzare il teorema centrale del limite e approssimare $Y \sim N(nE[X], nVar(X))$, da cui otteniamo (con la correzione di continuità)

$$P(Y < 100) \simeq P\left(Z < \frac{99.5 - 40 \cdot \frac{8}{3}}{\sqrt{40 \cdot \frac{8}{9}}}\right) = \Phi(-1.20) = 0.12.$$

Problema 2. Valentina, una ragazza molto appassionata di pallavolo, sta seguendo molto attentamente il campionato mondiale di pallavolo maschile che si sta svolgendo in Polonia. Durante la partita Italia-Belgio, il commentatore ha affermato che gli atleti del mondiale hanno un'elevazione verticale media maggiore di 80 cm. Valentina vuole impostare un test d'ipotesi per controllare che quanto detto dal commentatore sia vero, ma non è molto abile con la statistica e ha quindi bisogno del vostro aiuto!
Assumendo che le elevazioni verticali dei giocatori del mondiale siano distribuite come una normale di varianza 25,

a) impostare un test d'ipotesi per verificare l'affermazione del commentatore, specificando ipotesi nulla, ipotesi alternativa, regione critica di livello $\alpha = 0.01$.

Facendo una ricerca in internet, Valentina trova il valore del salto verticale di 16 giocatori di diverse nazionalità, ottenendo una media campionaria pari a 83.3 cm.

b) Qual è il risultato del test? Calcolare il p-value dei dati raccolti.

Supponendo che la reale media dell'elevazione verticale sia pari a 83 cm, calcolare

c) la probabilità di un errore di secondo tipo per il test utilizzato da Valentina;

d) il numero n di osservazioni con cui il test utilizzato da Valentina ha potenza maggiore di 0.8.

Risultati.

a) Sia X la variabile aleatoria "elevazione verticale di un giocatore del mondiale (in cm)". Possiamo assumere $X \sim N(\mu, 25)$. Detto $X_1, ..., X_n$ un campione di numerosità n, sia $\overline{X}_n$ la media campionaria. Il test da impostare è:
$$H_0 : \mu \leq 80 \qquad H_1 : \mu > 80,$$
con statistica test
$$Z_0 = \frac{\overline{X}_n - 80}{5/\sqrt{n}}$$
che sotto H_0 è distribuita come una normale standard. La regione critica di livello $\alpha = 0.01$ risulta quindi essere:
$$RC_{0.01} = \{Z_0 > z_{1-0.01}\} = \{Z_0 > 2.33\}.$$

b) Utilizzando i dati raccolti otteniamo $z_0 = \frac{83.3-80}{5/4} = 2.64 \in RC_{0.01}$ quindi rifiutiamo l'ipotesi nulla al livello $\alpha = 0.01$. Il p-value del test è
$$p = P(Z_0 > z_0) = P(Z_0 > 2.64) = 1 - \Phi(2.64) = 0.004.$$

c) Supponendo $\mu = 83$ e ponendo $Z = \dfrac{\overline{X}_n - 83}{5/\sqrt{n}}$, l'errore di secondo tipo del test è:
$$\beta = P_{\mu=83}(Z_0 \leq 2.33) = P_{\mu=83}\left(Z \leq 2.33 - \frac{3}{5/4}\right) = P_{\mu=83}(Z \leq -0.07) = \Phi(-0.07) = 0.472.$$

d) La potenza del test con un campione di dimensione n è data da
$$\gamma = P_{\mu=83}(Z_0 > 2.33) = P_{\mu=83}\left(Z > 2.33 - \frac{3}{5/\sqrt{n}}\right) = \Phi\left(-2.33 + \frac{3}{5/\sqrt{n}}\right).$$
Quindi da $\gamma > 0.80$ ricaviamo
$$-2.33 + \frac{3}{5/\sqrt{n}} > 0.85$$
da cui $\sqrt{n} > 5.3$. Sono quindi necessari almeno 29 valori, $n \geq 29$.

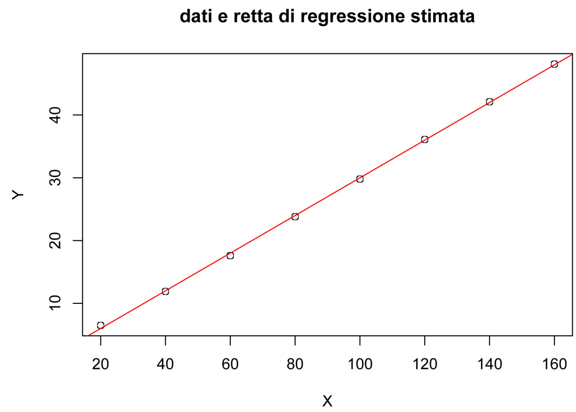
Problema 3. La legge di Hooke stabilisce che l'allungamento y subito da una molla è direttamente proporzionale alla forza x a essa applicata:

$$y = b_1 \, x, \tag{1}$$

dove la costante di proporzionalità b_1 viene detta costante elastica e dipende dalla natura del materiale di cui è costituita la molla.

Uno studente di ingegneria esegue alcune prove con 8 molle uguali, tutte di uno stesso materiale sconosciuto, ciascuna molla lunga 2 metri e di diametro pari a 0.34 mm, sottoponendole a diverse forze per stimare la costante b_1. Ecco i risultati e il relativo diagramma di dispersione:

forza	x (g)	20	40	60	80	100	120	140	160
allungamento	y (mm)	5.6	11.9	17.6	23.8	29.8	35.7	41.9	47.7



I dati raccolti vengono elaborati sulla base del modello empirico lineare e gaussiano

$$Y = b_0 + b_1 \, x + \epsilon, \qquad \epsilon \sim N(0; \sigma^2)$$

trovando i seguenti risultati

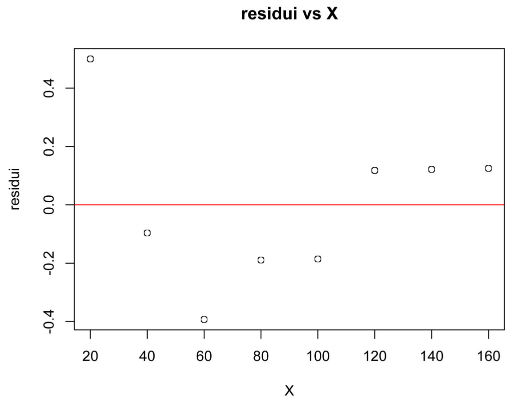
```
Residuals:
     Min       1Q   Median       3Q      Max
-0.39286 -0.18661  0.01071  0.12232  0.50000

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.003571   0.231193   0.015    0.988
X           0.299821   0.002289 130.975 1.34e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2967 on 6 degrees of freedom
Multiple R-squared: 0.9997,      Adjusted R-squared: 0.9996
F-statistic: 1.715e+04 on 1 and 6 DF,  p-value: 1.336e-11
------
Residuals


        Shapiro-Wilk normality test


data:  res
W = 0.9424, p-value = 0.6345
```



1. Quanto valgono le stime dei coefficienti b_0, b_1 e σ?
2. Valutare la bontà del modello empirico.
3. Fornire una stima intervallare al 95% per la costante elastica del materiale esaminato.
4. Sottoporre a verifica l'ipotesi che la costante elastica delle molle sia pari a 0.3. Si indichino le ipotesi statistiche da sottoporre a test, la regione critica di livello α, il p-value dei dati, la conclusione.
5. Sottoporre a verifica l'ipotesi di validità della legge di Hooke per il materiale esaminato, o, meglio, l'ipotesi che la legge di Hooke valga per l'allungamento atteso a forza fissata: $\mathbb{E}[Y] = b_1\, x$.

   Si indichino le ipotesi statistiche da sottoporre a test, la regione critica di livello α, il p-value dei dati, la conclusione.

Risultati.

1. $\widehat{b}_0 = 0.003571, \qquad \widehat{b}_1 = 0.299821, \qquad \widehat{\sigma} = 0.2967.$
2. Il modello empirico è buono: la percentuale di variabilità di y spiegata dalla regressione con la variabilità di x raggiunge il 99.97%, mentre la validità dell'ipotesi gaussiana è confermata dall'alto p-value del test di Shapiro-Wilk sui residui (0.6345) e dai grafici degli stessi (in realtà il grafico di dispersione evidenzia una possibile eteroschedasticità). Tuttavia solo il coefficiente angolare risulta significativo. Andrebbe confrontato con un modello senza intercetta.
3. L'intervallo di confidenza per b_1 è

$$\widehat{b}_1 - t_{n-2}\left(\frac{1+\gamma}{2}\right) \operatorname{se}(\widehat{b}_1) \leq b_1 \leq \widehat{b}_1 - t_{n-2}\left(\frac{1+\gamma}{2}\right) \operatorname{se}(\widehat{b}_1)$$

Poiché $t_6(0.975) = 2.447$ e $\operatorname{se}(\widehat{b}_1) = 0.002289$, otteniamo

$$0.299821 \pm 2.447 \cdot 0.00229 = 0.299821 \pm 0.0056 \implies IC(95\%) = (0.29422\ ,\ 0.30542).$$

4.

$$H_0\ :\ b_1 = 0.3 \qquad H_1\ :\ b_1 \neq 0.3$$
$$R_\alpha\ :\ \left|\widehat{b}_1 - 0.3\right| > \operatorname{se}(\widehat{b}_1)\, t_6(\alpha/2)$$

$$\frac{\left|\widehat{b}_1 - 0.3\right|}{\operatorname{se}(\widehat{b}_1)} = \frac{0.3 - 0.299821}{0.002289} = 0.0782 \implies 0.0782 = t_6(\alpha/2) < 0.265 = t_6(0.4) \implies \text{p-value} > 0.8$$

per cui, con un p-value superiore all'80% (valore esatto 94.02%), non possiamo rifiutare l'ipotesi nulla: la costante elastica del materiale esaminato è pari a 0.3.

5.

$$H_0\ :\ b_0 = 0 \qquad H_1\ :\ b_0 \neq 0$$
$$R_\alpha\ :\ \left|\widehat{b}_0\right| > \operatorname{se}(\widehat{b}_0)\, t_6(\alpha/2)$$
$$\text{p-value} = 0.988$$

per cui, con un p-value così alto, non possiamo rifiutare l'ipotesi nulla: vale la legge di Hooke per il materiale esaminato.