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

I Appello di Statistica per Ingegneria Fisica
21 luglio 2017

Cognome, Nome, Matricola:

Problema 1. Lo spara-ragnatele usato da Peter ha una gittata G gaussiana di media 34 metri e deviazione standard 5 metri.

1. Quale percentuale di ragnatele sparate ha una gittata superiore ai 40 metri?

Peter sta per sparare 67 ragnatele. Le gittate di più lanci sono indipendenti.

2. Quale è il numero atteso di volte che supererà i 40 metri?
3. Qual è probabilità che superi i 40 metri un numero di volte maggiore di 10?

Tony sta progettando un nuovo modello di spara-ragnatele per portare la percentuale calcolata in (1) al 30% almeno.

4. A parità di deviazione standard, quale gittata media ci vorrebbe?

Risultati.

1. $G \sim N(34\,,\,25)$ per cui

$$\mathbb{P}(G > 40) \;=\; \mathbb{P}\left(\frac{G-34}{5} > \frac{40-34}{5}\right) = 1 - \Phi(1.2) = 0.1151 = 11.51\%$$

2. Se X è il numero di ragnatele che superano i 40 metri su 67 lanci, allora $X \sim B(67, 0.1151)$ per cui

$$\mathbb{E}[X] = 67 \cdot 0.1151 = 7.71$$

3. $\mathbb{P}(X > 10) \;=\; \mathbb{P}(X \geq 11) \;=\; \sum_{k=11}^{67} \binom{67}{k} 0.1151^k\, 0.8849^{67-k} = 0.1432 = 14.32\%$

   Essendo $\mathbb{E}[X] = 7.71 > 5$ e $67 - \mathbb{E}[X] > 5$ vale anche il risultato approssimato
   $\mathbb{P}(X > 10) \;\simeq\; 1 - \Phi\left(\dfrac{10.5 - 7.71}{\sqrt{67 \cdot 0.1151 \cdot 0.8849}}\right) = 1 - \Phi(1.07) = 0.1427 = 14.27\%$

4. Se $G \sim N(\mu\,,\,25)$ allora

$$\mathbb{P}(G > 40) \;=\; \mathbb{P}\left(\frac{G-\mu}{5} > \frac{40-\mu}{5}\right) \;\geq\; 0.3$$

$$\frac{40-\mu}{5} \leq z_{0.3}$$

$$\mu \geq 40 - 5z_{0.3} = 40 - 5 \cdot 0.52 = 37.4$$

Problema 2. Peter deve testare i nuovi spara-ragnatele inviatigli dai Laboratori Stark. In particolare deve controllare che la gittata superi i 40 metri almeno il 30% delle volte. Non potendo sparare infinite ragnatele, Peter deve decidere sulla base di sole n ragnatele e chiede quindi il vostro aiuto, perché voi siete il suo "uomo sulla sedia".

1. Impostate un opportuno test statistico per provare con forte evidenza statistica che la gittata supera i 40 metri almeno il 30% delle volte. Specificate in particolare:
   - la variabile aleatoria di interesse, con la sua legge e il parametro incognito su cui inferire,
   - ipotesi nulla e ipotesi alternativa,
   - regione critica di livello α per decidere con un campione casuale di n ragnatele sparate,
   - eventuali condizioni affinchè il livello nominale approssimi bene quello reale.

Peter vuole limitare a 0.01 la probabilità di approvare uno spara-ragnatele malfunzionante. Allo stesso tempo vuole evitare la figuraccia di bocciare uno spara-ragnatele che in realtà rispetta gli standard dichiarati: per l'esattezza, vuole limitare a 0.4 la probabilità di una figuraccia nel caso in cui la gittata superi i 40 metri il 35% delle volte.

2. Quale livello α dovete scegliere per il test?
3. Quante ragnatele deve sparare Peter?

Alla fine Peter spara tutte le ragnatele possibili prima che il nuovo spara-ragnatele si scarichi: 67 ragnatele, di cui 29 superano i 40 metri.

4. Quanto vale il p-value dei dati raccolti?
5. Cosa può concludere Peter?

Risultati.

1.
   - Se G è la gittata di una ragnatela sparata dal nuovo dispositivo, allora la variabile di interesse è
$$X = \begin{cases} 1, & \text{se } G > 40, \\ 0, & \text{se } G \leq 40, \end{cases} \qquad \sim B(p),$$
dove p è la probabilità di avere $G > 40$, ovvero la proporzione di ragnatele con $G > 40$.

   - $H_0 : p < 0.3 \qquad \text{vs} \qquad H_1 : p \geq 0.3$

   - $R_\alpha : \widehat{p} = \frac{1}{n}\sum_{k=1}^{n} X_k > 0.3 + \sqrt{\frac{0.3 \cdot 0.7}{n}}\, z_\alpha$

   - Le gittate di più lanci devono essere indipendenti, ed inoltre ci vuole $0.3 \cdot n > 5 \quad$ e $\quad 0.7 \cdot n > 5$

2. $\alpha = 0.01$

3. Supponendo $\quad 0.35 \cdot n > 5 \quad$ e $\quad 0.65 \cdot n > 5,$

$$\beta = \mathbb{P}_{p=0.35}\left(\widehat{p} \leq 0.3 + \sqrt{\frac{0.3 \cdot 0.7}{n}}\, z_{0.01}\right) = \mathbb{P}_{p=0.35}\left(\frac{\widehat{p} - 0.35}{\sqrt{\frac{0.35 \cdot 0.65}{n}}} \leq \frac{-0.05 + \sqrt{\frac{0.3 \cdot 0.7}{n}}\, z_{0.01}}{\sqrt{\frac{0.35 \cdot 0.65}{n}}}\right)$$
$$\simeq \Phi\left(-\frac{0.05}{\sqrt{0.35 \cdot 0.65}}\sqrt{n} + \sqrt{\frac{0.3 \cdot 0.7}{0.35 \cdot 0.65}}\, z_{0.01}\right) < 0.4$$

$$-\frac{0.05}{\sqrt{0.35 \cdot 0.65}}\sqrt{n} + \sqrt{\frac{0.3 \cdot 0.7}{0.35 \cdot 0.65}}\, z_{0.01} < -z_{0.4}$$

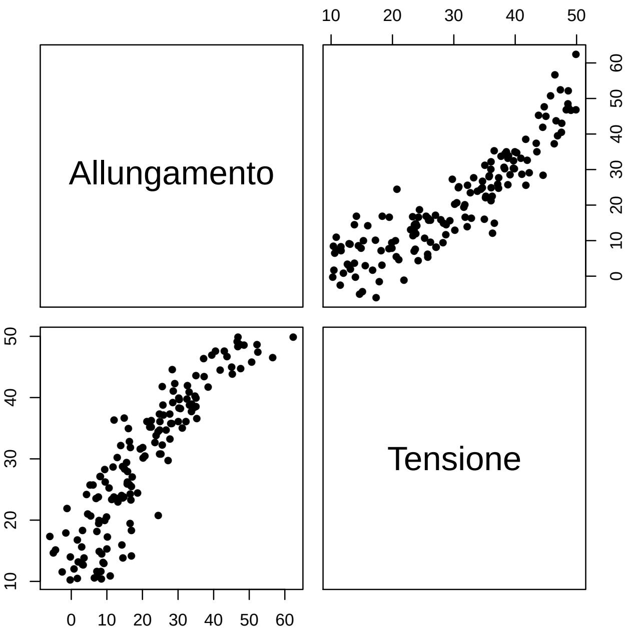
$$n \geq 564$$

che è coerente con la supposizione iniziale $\quad 0.35 \cdot n > 5 \quad$ e $\quad 0.65 \cdot n > 5.$

4. I dati raccolti danno $\widehat{p} = \frac{29}{67} = 0.4328$ e un p-value α
$$\frac{29}{67} = 0.3 + \sqrt{\frac{0.3 \cdot 0.7}{67}}\, z_\alpha$$
$$\alpha = 1 - \Phi(z_\alpha) = 1 - \Phi(2.37) = 0.0089$$

5. Essendo il p-value pari a $0.0089 < 0.01$, Peter può rifiutare H_0 ad un livello $\alpha = 0.01$: c'è forte evidenza statistica che la gittata del nuovo spara-ragnatele superi i 40 metri almeno il 30% delle volte.

Problema 3. Le ragnatele di Peter sono costituite da un materiale elastico molto resistente che, quando viene sollecitato con una certa tensione, si deforma allungandosi. Con la supervisione di Tony, Peter decide di studiare le caratteristiche elastiche delle proprie ragnatele. A questo scopo effettua una serie di prove minuziose in cui impone una tensione t (misurata in Kg) su uno stesso filo di ragnatela, e misura l'allungamento conseguente A (in mm) ottenendo i seguenti dati.



Peter ritiene che, mediando su fattori secondari non in esame, il risultante allungamento medio debba soddisfare una delle relazioni seguenti:

E_1: $\mathbb{E}[A|t] = \beta_0 + \beta_1 \cdot t$,

E_2: $\mathbb{E}[A|t] = \beta_0 + \beta_1 \cdot t^2$.

Elabora pertanto i dati raccolti sulla base dei due modelli di regressione empirici gaussiani M_1 ed M_2 associati alle relazioni E_1 ed E_2. Le tabelle riassuntive delle regressioni e i grafici di diagnostica sono riportati di seguito. I p-value dei test Shapiro-Wilks dei residui valgono, rispettivamente, 0.7637 e 0.9863.

1. Specificare i modelli M_1 ed M_2 impiegati da Peter.

2. Fornire una stima delle relazioni E_1 ed E_2 a partire dai dati.

3. Quale relazione sembra essere meglio supportata dai dati? Motivare la risposta in base a tutte le informazioni a disposizione.

4. A partire dalla relazione scelta, quale allungamento medio ci si deve aspettare in corrispondenza del valore $t = 0$? A quale significato fisico corrisponde ciò?

5. Quante prove ha sostenuto Peter? Cosa dicono i grafici della prova 108?

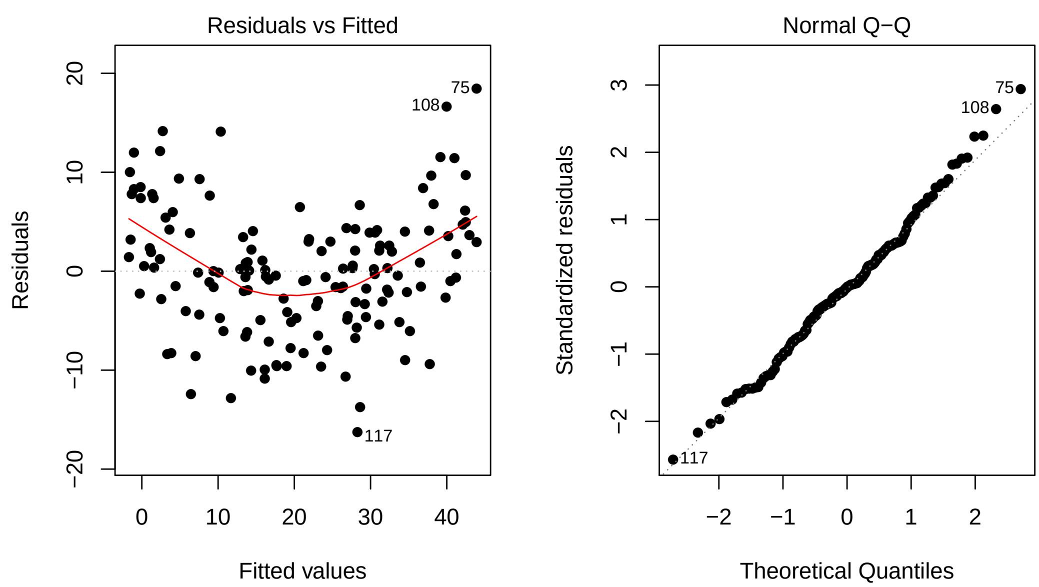
6. Sapendo che per il dataset raccolto da Peter $\bar{t} = 29.67$ Kg, $\overline{t^2} = 940.25$ Kg, $S_t^2 = 128.9039$ Kg^2, $S_{t^2}^2 = 401990$ Kg^4, e che Peter pesa 70 Kg, fornire una previsione intervallare al 95% per l'allungamento quando userà le ragnatele durante la sua prossima missione. Tale previsione è affidabile?

```
Call:
lm(formula = Allungamento ~ Tensione, data = Data)

Residuals:
     Min       1Q   Median       3Q      Max 
-16.2426  -4.3052  -0.0599   3.8763  18.4432 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -13.51540    1.45749  -9.273   <2e-16 ***
Tensione      1.15088    0.04589  25.078   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.361 on 148 degrees of freedom
Multiple R-squared:  0.8095,    Adjusted R-squared:  0.8082 
F-statistic: 628.9 on 1 and 148 DF,  p-value: < 2.2e-16
```

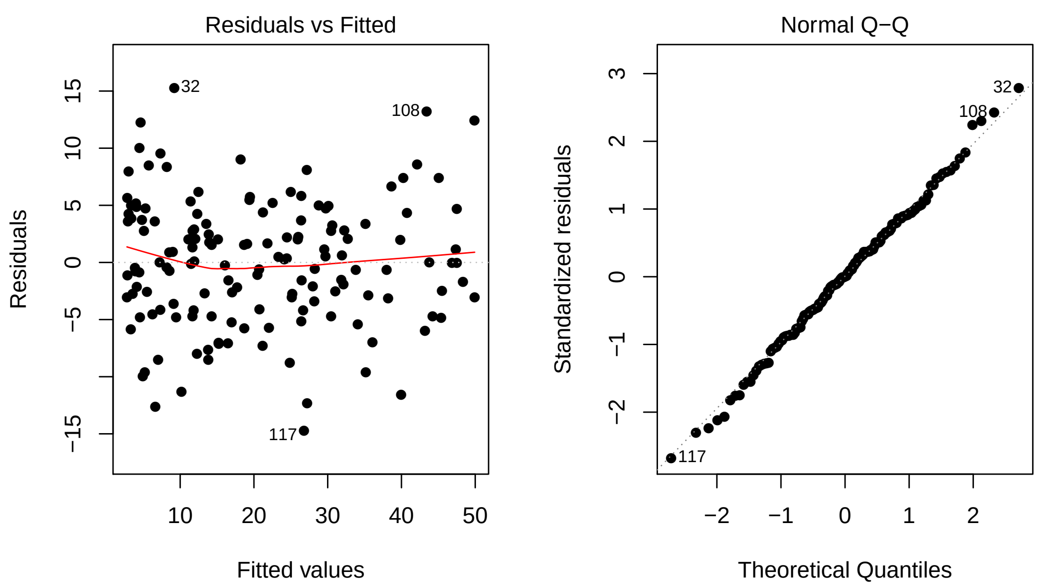


```
Call:
lm(formula = Allungamento ~ I(Tensione^2), data = Data)

Residuals:
     Min       1Q   Median       3Q      Max
-14.7339  -3.5666   0.0062   3.6649  15.2774

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.6979121  0.8078622   0.864    0.389
I(Tensione^2) 0.0197671  0.0006648  29.736   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.519 on 148 degrees of freedom
Multiple R-squared:  0.8566,    Adjusted R-squared:  0.8557
F-statistic: 884.2 on 1 and 148 DF,  p-value: < 2.2e-16
```



Risultati.

1. M_1: $A = \beta_0 + \beta_1 \cdot t + \varepsilon$, con $\varepsilon \sim \mathcal{N}(0, \sigma^2)$,

   M_2: $A = \beta_0 + \beta_1 \cdot t^2 + \varepsilon$, con $\varepsilon \sim \mathcal{N}(0, \sigma^2)$.

2. Secondo M_1, per E_1 si ha $\widehat{\beta}_0 = -13.51$ e $\widehat{\beta}_1 = 1.15$, mentre secondo M_2 per E_2 si ha $\widehat{\beta}_0 = 0.6979$ e $\widehat{\beta}_1 = 0.0197$.

3. La relazione meglio supportata dai dati è E_2, in cui la dipendenza quadratica da t coglie meglio l'andamento di A. Il modello M_2 è infatti preferibile ad M_1 sia per il maggior valore di R^2, sia per i residui, gaussiani ed omoschedastici, mentre per M_1 si vede un chiaro pattern parabolico, segno di una non-linearità nei dati non colta nel modello.

   Il modello M_2 ha un'intercetta poco significativa e se ne potrebbe valutare l'eliminazione, visto il suo significato fisico (vedi punto seguente).

4. Il modello M_2 fornisce una stima puntuale di $\widehat{A}(0) = 0.69$ mm, ma essendo l'intercetta poco significativa ($p = 0.389$), sembra possibile considerare un valore nullo. Effettivmente, in mancanza di sollecitazione, non ci attendiamo alcun allungamento medio della ragnatela.

5. Dalle tabelle riassuntive si deduce che ha sostenuto $n = 148 + 2$ prove. Il punto di indice 108 è outlier, come si deduce dal grafico di diagnostica.

6. L'intervallo di previsione, al 95%, vale $IP_{0.95} = [85.27, 109.84] = 97.56 \pm 12.28$ mm.

   La previsione non è affidabile: anche se il modello è molto buono, $t = 70$ Kg è lontano dal range dei dati su cui il modello è stato costruito.