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

III Appello di Statistica per Ingegneria Fisica
23 febbraio 2018

Cognome, Nome, Matricola:

Problema 1. Il nuovo misuratore di resistenza introduce un errore casuale X che, fissate opportunamente le unità di misura, ha densità continua della forma

$$f_X(t) = k\, t^2\, I_{[-1,1]}(t),$$

dove k è un numero reale.

1. Tracciare un grafico qualitativo di f.
2. Calcolare k.
3. Calcolare l'errore atteso $\mathbb{E}[X]$.
4. Calcolare la varianza dell'errore $\mathrm{Var}[X]$.

Sono giudicati accettabili gli errori con $|X| < \dfrac{1}{10}$.

5. Calcolare $\mathbb{P}\left(\left|X\right| < \dfrac{1}{10}\right)$.

Insoddisfatti del risultato, ci vediamo costretti ad eseguire n misure indipendenti per ridurre l'errore di misura. Sia quindi $\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$ l'errore di misura per la media campionaria di n misure di errori indipendenti $X_1, \ldots, X_n$.

6. Trovare quante misure servono per avre $\mathbb{P}\left(\left|\overline{X}_n\right| < \dfrac{1}{10}\right) > 0.8$.

Risultati.

2. Poichè
$$1 = \int_{-1}^{1} ks^2\, ds = \frac{2k}{3},$$
si ottiene $k = 3/2$.

3. Dalla simmetria della densità (o anche integrando $\int_{-1}^{1} 3/2 s^3\, ds$) si ottiene $\mathbb{E}[X] = 0$.

4. Per il punto precedente vale $\mathrm{Var}[X] = \mathbb{E}[X^2]$, dunque
$$\mathrm{Var}[X] = \int_{-1}^{1} \frac{3}{2} s^4\, ds = \frac{3}{5}.$$

5.
$$P\Big(-\frac{1}{10} < X < \frac{1}{10}\Big) = \int_{-1/10}^{1/10} \frac{3}{2} s^2\, ds = \frac{1}{1000}.$$

6. Per il teorema centrale del limite, se n è grande, la variabile $\bar{X}_n$ ha distribuzione approssimativamente normale con media e varianza rispettivamente
$$\mathbb{E}[\bar{X}_n] = 0, \qquad \mathrm{Var}[\bar{X}_n] = \frac{3}{5} \cdot \frac{1}{n};$$
dunque se $Z \sim N(0,1)$ vale che
$$\begin{aligned} 0.8 < P\Big(-\frac{1}{10} < \bar{X}_n < \frac{1}{10}\Big) &= P\Big(-\frac{1/10}{\sqrt{3/5n}} < Z < \frac{1/10}{\sqrt{3/5n}}\Big) = \\ &= 2\Phi\left(\sqrt{5n/300}\right) - 1 \end{aligned}$$
da cui
$$\Phi\left(\sqrt{5n/300}\right) > 0.9 \Longrightarrow \sqrt{5n/300} > 1.282,$$
cioè $n > 98.6$, che implica $n \geq 99$ (coerente con l'ipotesi che sia grande).

Problema 2. L'Imperatore vuole stimare la proporzione p di sudditi favorevoli all'organizzazione della III Adunata. A tal fine ordina di intervistare un campione casuale di n sudditi.

(a) Si stabilisca quale debba essere la numerosità minima n_0 affinché una stima intervallare bilatera per p di livello 90% abbia ampiezza inferiore a 0.1.

Eseguita l'indagine, 203 degli n_0 sudditi intervistati si sono dichiarati favorevoli alla III Adunata.

(b) Si forniscano una stima puntuale e una stima intervallare di livello 90% per p.

Si supponga ora di sapere che $p = 0.8$.

(c) Calcolare in modo esatto la probabilità che un'indagine condotta su solo $n = 10$ sudditi fornisca un errore di stima puntuale di p inferiore a 0.05.

Risultati. Sia $\bar{x}$ la frazione di intervistati favorevole alla III Adunata.

(a) $A = 2\sqrt{\dfrac{\bar{x}(1-\bar{x})}{n}}\, z_{0.05} \leq A_0 = \dfrac{z_{0.05}}{\sqrt{n}}$, per cui $A \leq 0.1$ se $A_0 \leq 0.1$, che accade se $n \geq \left(\dfrac{z_{0.05}}{0.1}\right)^2 = 270.6$. Pertanto $n_0 = 271$.

(b) Stima puntuale: $\bar{x} = 203/271 = 0.7491$.

Stima intervallare: $\bar{x} \pm \sqrt{\frac{\bar{x}(1-\bar{x})}{n_0}}\, z_{0.05} = 0.7491 \pm 0.0433$.

(c) $\mathbb{P}\Big(|0.8 - \overline{X}| < 0.05\Big) = \mathbb{P}\Big(|8 - Y| < 0.5\Big) = \mathbb{P}\Big(Y = 8\Big) = 0.3020$, dove

$Y =$ numero di favorevoli alla III Adunata su 10 sudditi intervistati $\sim B(10, 0.8)$.

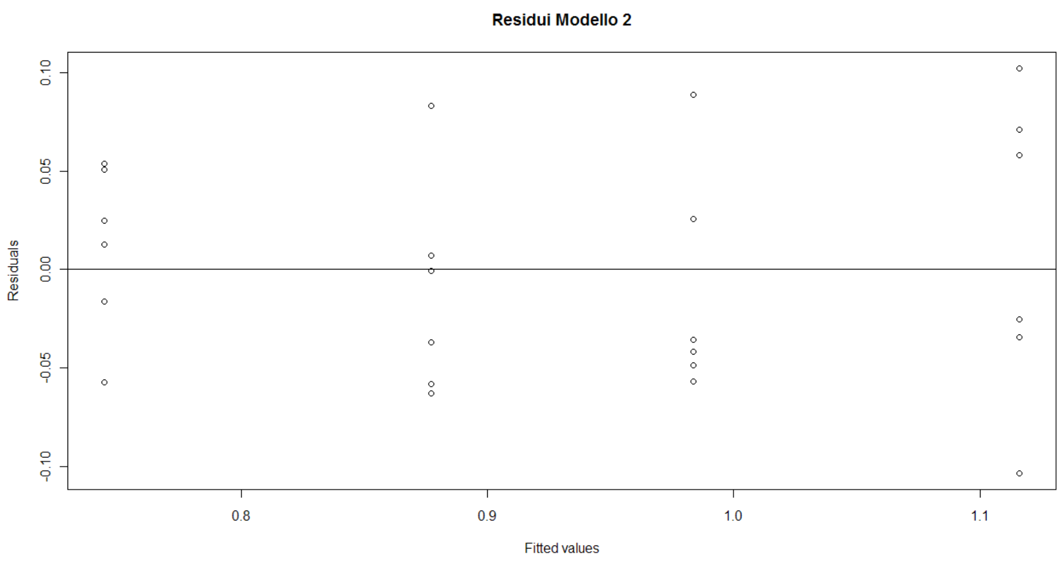
Problema 3. La società Firebolt s.r.l ha elaborato un modello di regressione lineare empirico gaussiano per il suo nuovo saldatore che lega il rapporto di forma H (cioè il rapporto fra profondità e larghezza di un cordone saldato) alla potenza dell'impulso P (in kW) e al diametro di uno spot D (in mm).

1. Scrivere la relazione fra H, P e D ipotizzata da un modello di regressione di H su P e D.
2. Quanto vale la variazione media di H se, a parità di impulso P, il diamertro D aumenta di 0.1 mm?

Elaborati con R i dati sperimentali raccolti, sono stati ottenuti i risultati mostrati di seguito (valori di sintesi, grafico dei residui, p-value del test di Shapiro-Wilk per i residui).

3. Quanti casi sono stati analizzati per elaborare il modello?
4. Quale percentuale della variabilità di H è spiegata dal modello elaborato?
5. È possibile usare questo modello per fare inferenza e previsioni intervallari. Perché?
6. Dare una stima puntuale della variazione media di H per un aumento di 0.1 mm di D, a parità di P.
7. Dare una stima intervallare di livello 95% della variazione media di H per un aumento di 0.1 mm di D, a parità di P.

```
> summary(Model2)

Call:
lm(formula = H ~ P + D)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.103125 -0.043292 -0.008583  0.051458  0.101875 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.53096    0.07149   7.427 2.65e-07 ***
P            0.47850    0.04834   9.899 2.31e-09 ***
D           -0.44194    0.08056  -5.486 1.92e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0592 on 21 degrees of freedom
Multiple R-squared: 0.8591,     Adjusted R-squared: 0.8457 
F-statistic: 64.05 on 2 and 21 DF,  p-value: 1.153e-09

> shapiro.test(Model2$residuals)

        Shapiro-Wilk normality test

data:  Model2$residuals 
W = 0.9519, p-value = 0.2975
```



Risultati.

1. $H = \beta_0 + \beta_1 P + \beta_2 D + \epsilon$, dove $\epsilon \sim N(0, \sigma^2)$.

2. $\mathbb{E}[H|P, D + 0.1] - \mathbb{E}[H|P, D] = \beta_2 \times 0.1$

3. $n = 24$.

4. 85.91%.

5. Perché l'ipotesi gaussiana può essere accettata: p-value di Shapiro Wiks abbastanza alto e grafico dei residui buono (potrebbe evidenziare una qualche eteroschedasticità, ma non in modo preoccupante).

6. $\widehat{H}|_{P,D+0.1} - \widehat{H}|_{P,D} = \widehat{\beta}_2 \times 0.1 = -0.44194 \times 0.1 = -0.044194$

   quindi stimiamo che, se il diametro dello spot D aumenta di 1 mm e P non varia, il rapporto di forma H in media diminuisce di 0.044194.

7. Un intervallo di confidenza bilatero al 95% per β_2 è

   $$\widehat{\beta}_2 \pm \mathrm{se}(\widehat{\beta}_2)\, t_{0.025}(21) = -0.44194 \pm 0.08056 \times 2.080 = -0.44194 \pm 0.16756 = (-0.60950\,,\, -0.27438)$$

   per cui un intervallo di confidenza bilatero al 95% per $0.1 \times \beta_2$ è

   $$(-0.060950\,,\, -0.027438)$$