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

I APPELLO DI STATISTICA PER INGEGNERIA ENERGETICA
25 febbraio 2015

Cognome, Nome e Numero di matricola:

Problema 1. A Mos Eisley le navi-cargo decollano secondo un processo di Poisson, al ritmo di 6 all'ora.

(a) Quanto tempo passa mediamente fra un decollo e il successivo?

(b) Trovare il minimo tempo di attesa t_0 per assistere ad almeno un decollo con una probabilità maggiore o uguale al 99.9%.

(c) Quante navi-cargo decollano mediamente in un tempo t_0?

(d) Calcolare la probabilità che in t_0 decollino almeno 3 navi cargo.

(e) Calcolare la probabilità che in t_0 decollino almeno 9 navi cargo.

Risultati. Sia N_t il numero di decolli in un tempo t, per cui $N_t \sim P(\nu t)$. Misurando il tempo in ore, $\nu = 6$.

(a) Sia T il tempo di attesa fra due decolli. Allora $\mathbb{E}[T] = 1/\nu = 1/6$ ore $= 10'$.

(b)
$$0.999 \le \mathbb{P}(N_t \ge 1) = \mathbb{P}(T \le t) = 1 - \mathrm{e}^{-\nu t}$$
$$\mathrm{e}^{-\nu t} \le 0.001$$
$$-\nu t \le \ln 0.001$$
$$\nu t \ge \ln 1000$$
$$t \ge \frac{\ln 1000}{\nu} = 1.151292546 \text{ ore} = 69.07755279 \text{ min} = 1 \text{ ora } 9'5'' = t_0$$

(c) $\mathbb{E}[N_{t_0}] = \nu\, t_0 = \ln 1000 = 6.91$.

(d) $\mathbb{P}(N_{t_0} \ge 3) = 1 - \mathrm{e}^{-\nu\, t_0}\left(1 + \nu\, t_0 + \dfrac{(\nu\, t_0)^2}{2}\right) = 0.9682.$

(e) Poiché $\nu\, t_0 > 5$, anche se non di molto, ricorriamo all'approssimazione normale, trovando
$$\mathbb{P}(N_{t_0} \ge 9) = \mathbb{P}\left(\frac{N_{t_0} - \nu\, t_0}{\sqrt{\nu\, t_0}} \ge \frac{8.5 - \nu\, t_0}{\sqrt{\nu\, t_0}}\right) \simeq 1 - \Phi(0.61) = 0.27.$$
Il valore esatto invece è $\mathbb{P}(N_{t_0} \ge 9) = 0.26$.

Problema 2. Il diametro X dei cilindretti prodotti dalla H&A deve essere rigorosamento compreso fra 20.99 cm e 21.01 cm. La H&A tollera che le specifiche siano violate al massimo dal 5% dei cilindretti, e solo in questo caso ritiene il processo produttivo *sotto controllo* e inutile una revisione. Per questo motivo ha appena controllato i diametri di un campione casuale di 170 cilindretti, trovando

$$\overline{x}_{170} = 20,99988464 \text{ cm}, \qquad s_{170} = 0,005571133 \text{ cm},$$

e la seguente tabella di distribuzione di frequenza

Classi	Frequenze assolute
[20.988; 20.990)	5
[20.990; 20.992)	10
[20.992; 20.994)	16
[20.994; 20.996)	19
[20.996; 20.998)	16
[20.998; 21.000)	21
[21.000; 21.002)	19
[21.002; 21.004)	25
[21.004; 21.006)	10
[21.006; 21.008)	13
[21.008; 21.010)	13
[21.010; 21.012)	3

(a) Rappresentare la distribuzione riportata in tabella mediante un istogramma.

(b) Stimare la percentuale di cilindretti che violano le specifiche, sia mediante una stima puntuale, sia mediante una stima intervallare di livello di confidenza 0.95.

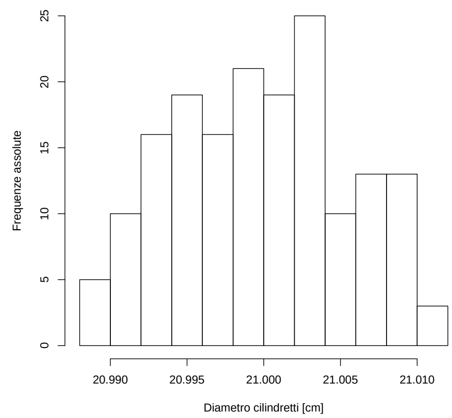
Ancora indecisa circa l'opportunità di una revisione, la H&A stabilisce che si può evitare la revisione solo se i dati raccolti provano con forte evidenza statistica che il processo produttivo è *sotto controllo*.

(c) Impostare un opportuno test di ipotesi per decidere se effettuare una revisione. Specificare ipotesi nulla, ipotesi alternativa, regione critica ed eventuali condizioni di utilizzo del test.

(d) Calcolare il p-value dei dati raccolti per il test appena impostato e trarre le debite conclusioni circa la necessità di una revisione.

Risultati.

(a)



(b) Sia Y la variabile bernoulliana che indica se $|X - 21| > 0.01$.

Stima puntuale di $p = \mathbb{P}\Big(|X - 21| > 0.01\Big) = \mathbb{P}(Y = 1)$:

$$\widehat{p} = \overline{y}_{170} = \frac{8}{170} = 0.0471.$$

Stima intervallare di $p = \mathbb{P}\Big(|X - 21| > 0.01\Big) = \mathbb{P}(Y = 1)$ di livello di confidenza 0.95:

$$\widehat{p} \pm \sqrt{\frac{\widehat{p}(1-\widehat{p})}{170}}\, z_{0.025} = 0.0471 \pm 0.0319 = (0.0152\ ,\ 0.0789).$$

(c) Mettiamo a verifica le ipotesi $\quad H_0 : p > 0.05 \quad$ contro $\quad H_1 : p \leq 0.05$.

Poiché sia $n\,\overline{y}_{170} = 8 > 5$, sia $n\,(1 - \overline{y}_{170}) = 162 > 5$, possiamo usare la regione critica di livello α

$$R_\alpha : \widehat{p} < 0.05 - \sqrt{\frac{0.05 \cdot 0.95}{170}}\, z_\alpha$$

(d) Il p-value dei dati raccolti è il valore di α soluzione di

$$\widehat{p} = 0.05 - \sqrt{\frac{0.05 \cdot 0.95}{170}}\, z_\alpha.$$

$$z_\alpha = \frac{(0.05 - \widehat{p})\sqrt{170}}{\sqrt{0.05 \cdot 0.95}}$$

$$\text{p-value} = \alpha = 1 - \Phi\left(\frac{(0.05 - \widehat{p})\sqrt{170}}{\sqrt{0.05 \cdot 0.95}}\right) = 1 - \Phi(0.176) = 0.4302$$

Nonostante sia $\widehat{p} \leq 0.05$, questo non basta per affermare con forte evidenza statistica che $p \leq 0.05$: i dati hanno un p-value del 43%, ben al di sopra degli usuali livelli di significatività, per cui non ci consentono di rifiutare l'ipotesi nulla $p \geq 0.05$. Revisione sia.

Problema 3. Il Dottor H. sta sviluppando un modello lineare empirico gaussiano che spieghi l'energia Y consumata al minuto da un automa MM con il suo peso x. Pertanto ordina al Barone A. di raccogliere i dati relativi a 21 differenti automi MM, per poi elaborarli con una regressione lineare semplice di Y su x. In allegato trovate lo specchietto riassuntivo della regressione, alcuni grafici dei residui standardizzati, il p-value dei residui standardizzati per il test di normalità di Shapiro-Wilk.

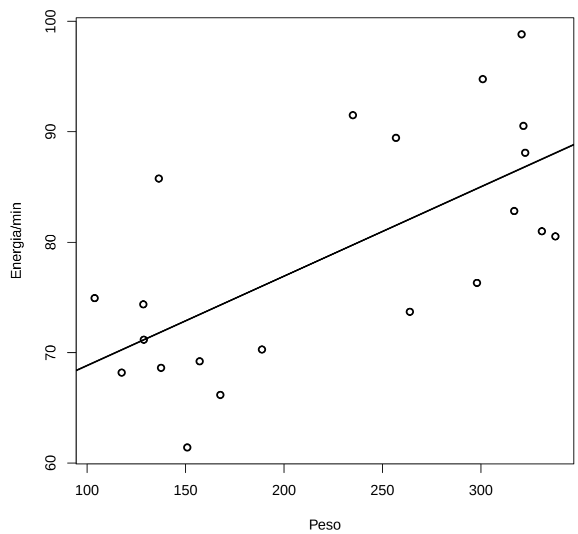
1. Scrivere il legame fra Y e x ipotizzato dal Dottor H.

2. Valutare la bontà dell'ipotesi gaussiana.

3. Sapendo che per il terzo automa MM è stato registrato un peso $x_3 = 150.86$ ed un'energia al minuto $y_3 = 61.42$, calcolare il corrispondente residuo standardizzato.

4. Trovare $\overline{x}$ ed S_{xx}, ovvero la media aritmetica dei 21 pesi registrati dal Barone A. e la somma dei loro scarti quadratici da tale media.

5. Prevedere con un intervallo al 95% l'energia Y consumata al minuto da un automa MM di peso $x = 350$.

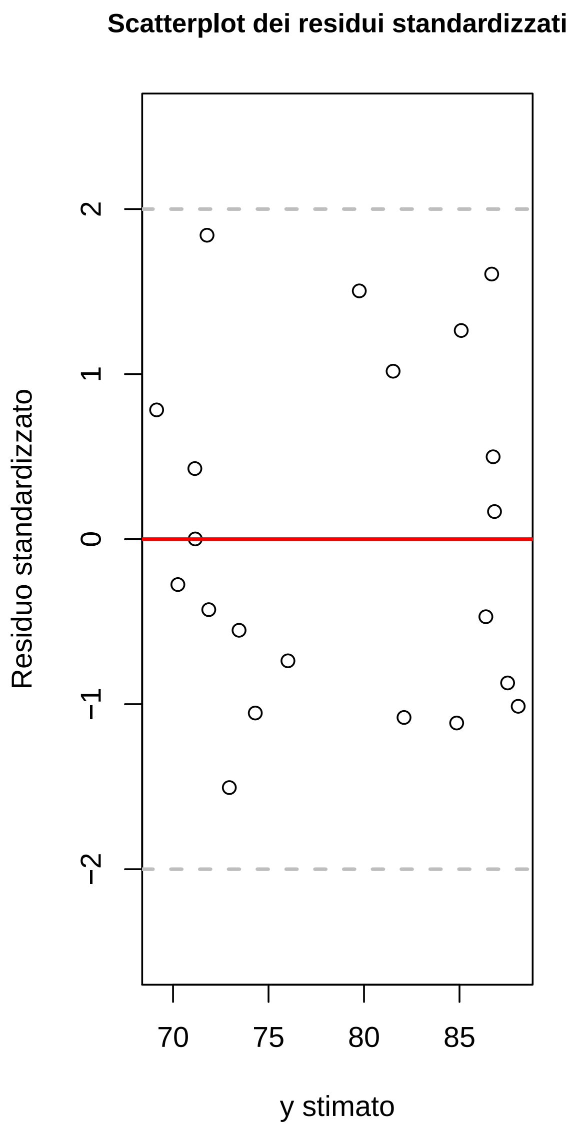


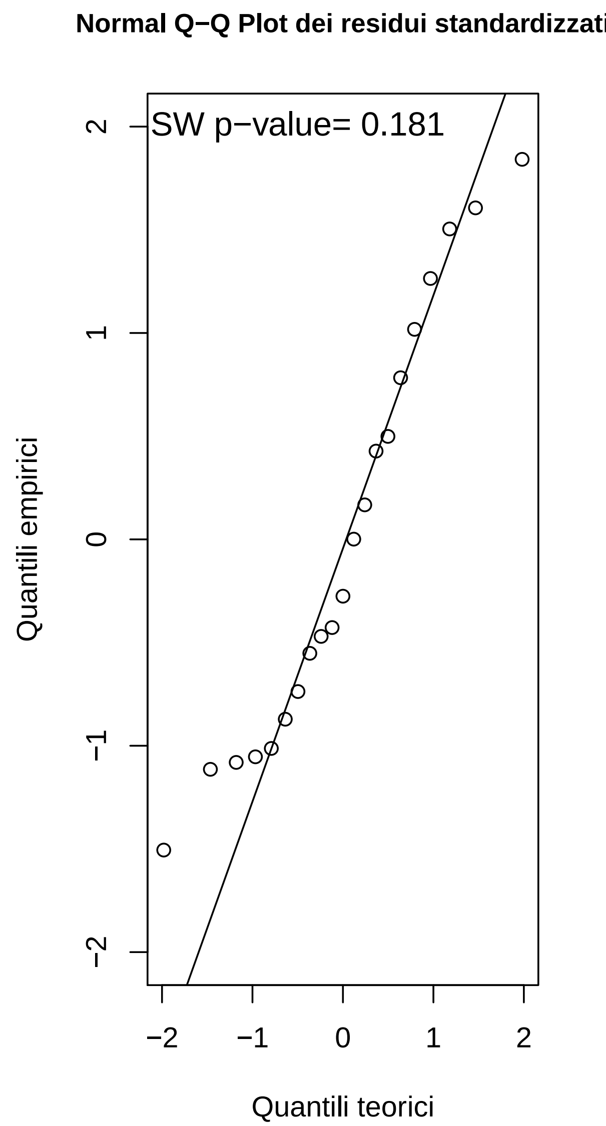

```
Call:
lm(formula = y ~ x, data = dati)

Residuals:
    Min      1Q  Median      3Q     Max 
-11.530  -6.531  -2.061   5.793  13.983 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 60.74030    4.97835  12.201 1.96e-10 ***
x            0.08092    0.02073   3.903 0.000956 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.004 on 19 degrees of freedom
Multiple R-squared:  0.445,	Adjusted R-squared:  0.4158 
F-statistic: 15.23 on 1 and 19 DF,  p-value: 0.0009562
```


Scatterplot dei residui standardizzati
Residuo standardizzato
y stimato
70
75
80
85
2
1
0
−1
−2


Normal Q−Q Plot dei residui standardizzati
SW p−value= 0.181
Quantili empirici
Quantili teorici
−2
−1
0
1
2

Risultati.

1. $Y = \beta_0 + \beta_1\, x + \epsilon$, dove $\epsilon \sim N(0, \sigma^2)$.

2. Ipotesi gaussiana non può essere rifiutata: residui omoschedastici e p-value di Shapiro-Wilk (0.18) non altissimo, ma comunque superiore al 10%.

3. $r_3 = \dfrac{y_3 - \widehat{\beta}_0 - \widehat{\beta}_1\, x_3}{\widehat{\sigma}} = -1.48.$

4. Possiamo ricavare $\overline{x}$ dagli errori standard degli stimatori $\widehat{\beta}_0$ e $\widehat{\beta}_1$:

$$\mathrm{se}(\widehat{\beta}_1) = \sqrt{\frac{\widehat{\sigma}^2}{S_{xx}}} \implies S_{xx} = \frac{\widehat{\sigma}^2}{\mathrm{se}^2(\widehat{\beta}_1)} = 149051.0623$$

$$\mathrm{se}(\widehat{\beta}_0) = \sqrt{\widehat{\sigma}^2\left(\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}}\right)} \implies \overline{x} = \sqrt{\left(\frac{\mathrm{se}^2(\widehat{\beta}_0)}{\widehat{\sigma}^2} - \frac{1}{n}\right)\frac{\widehat{\sigma}^2}{\mathrm{se}^2(\widehat{\beta}_1)}} = 224.86$$

5. Grazie alla bontà dell'ipotesi gaussiana e alla vicinanza del peso 350 ai pesi usati per elaborare il modello, possiamo prevedere ad un livello 95% che l'energia Y consumata al minuto da un automa MM di peso $x = 350$ sarà compresa nell'intervallo

$$\widehat{\beta}_0 + \widehat{\beta}_1\, 350 \pm \widehat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(350-\overline{x})^2}{S_{xx}}}\; t_{0.025}(n-2) = 89.06 \pm 18.78 = (70.28\ ,\ 107.85).$$