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

APPELLO DI STATISTICA PER INGEGNERIA ENERGETICA
12 settembre 2014

Cognome, Nome e Numero di matricola:

Problema 1. Il numero di chiamate alla centrale del 118 di una paese di piccole dimensioni si può modellizzare come un processo di Poisson di intensità $\nu = 3$ chiamate/ora. Domani Nino sarà di turno al centralino dalle 20:00 alle 24:00. Calcolare la probabilità che Nino riceva

(a) al massimo 3 chiamate,

(b) al massimo 15 chiamate,

(c) la seconda telefonata dopo le 20:30,

(d) la prima telefonata entro le 20:12.

Domani Ugo sarà di turno al centralino dalle 16:00 alle 20:00. Calcolare la probabilità che

(e) Nino e Ugo ricevano, in totale, al massimo 30 chiamate,

(facoltativo) Nino e Ugo ricevano, ciascuno, al massimo 15 chiamate.

Risultati. Chiamiamo N_t il processo di conteggio delle chiamate alla centrale del 118 durante il turno di Nino. Sappiamo che $N_t \sim \mathcal{P}(3t)$ dove il tempo t è espresso in ore. Pertanto:

(a)

$$\mathbb{P}(N_4 \leq 3) = \mathrm{e}^{-12}\left[1 + 12 + \frac{12^2}{2} + \frac{12^3}{3!}\right] = 0,0023.$$

(b)

$$\mathbb{P}(N_4 \leq 15) = \sum_{k=0}^{15} \mathrm{e}^{-12}\frac{12^k}{k!} = 0.8444.$$

oppure, con approssimazione normale ($12 > 5$),

$$\mathbb{P}(N_4 \leq 15) \simeq \Phi\left(\frac{15.5 - 12}{\sqrt{12}}\right) = \Phi(1.01) = 0.8438.$$

(c)

$$\mathbb{P}(N_{0,5} \leq 1) = \mathrm{e}^{-1,5}[1 + 1,5] = 0,5578.$$

(d) Sia T l'istante (dall'inizio del turno) della prima chiamata. Sappiamo che $T \sim Exp(3)$, e dunque

$$\mathbb{P}(T \leq 0.2) = \mathbb{P}(N_{0.2} \geq 1) = 1 - \mathrm{e}^{-3\cdot 0,2} = 0,4512.$$

Chiamiamo X_t il processo di conteggio delle chiamate alla centrale del 118 a partire dal turno di Ugo, dalle 16:00. Sappiamo che $X_t \sim \mathcal{P}(3t)$ dove il tempo t è espresso in ore. Pertanto:

(e)

$$\mathbb{P}(X_8 \leq 30) = 0.9042 \simeq \Phi\left(\frac{30.5 - 24}{\sqrt{24}}\right) = \Phi(1.33) = 0.9077.$$

(f) Grazie all'indipendenza degli incrementi di un processo di Poisson,

$$\begin{aligned}\mathbb{P}(X_4 \leq 15, X_8 - X_4 \leq 15) = \mathbb{P}(X_4 \leq 15)\,\mathbb{P}(X_8 - X_4 \leq 15) &= \mathbb{P}(X_4 \leq 15)\,\mathbb{P}(N_4 \leq 15)\\ &= \mathbb{P}(N_4 \leq 15)^2 = 0.7130 \simeq 0.7121.\end{aligned}$$

Problema 2. Alessandro deve acquistare un macchinario riempi sacchi di farina. Nando e Pino gli propongono ciascuno il proprio macchinario. Entrambi riempiono i sacchi con un contenuto medio di 5 kg di farina, ma lo fanno con una precisione che potrebbe essere diversa. Nando sostiene che il suo macchinario, seppur più costoso, è sicuramente più preciso. Pino sostiene che il suo macchinario, non solo è meno costoso, ma ha sicuramente la stessa precisione del macchinario di Nando. Per decidere quale credito dare a queste affermazioni, Alessandro verifica di persona la precisione dei due macchinari: 61 sacchi riempiti col macchinario di Nando danno una media e una varianza campionaria pari a

$$\overline{x}_{61} = 5.04\,\mathrm{kg}, \qquad s_x^2 = 0.0862\,\mathrm{kg}^2,$$

mentre 41 sacchi riempiti col macchinario di Pino danno una media e una varianza campionaria pari a

$$\overline{y}_{41} = 5.07\,\mathrm{kg}, \qquad s_y^2 = 0.0885\,\mathrm{kg}^2.$$

Nando pensa di aver concluso già l'affare, ma Alessandro non si accontenta di sole stime puntuali...
Si verifichi innanzitutto l'affermazione di Nando.

1. Si imposti un opportuno test statistico, specificando variabili aleatorie di interesse, ipotesi statistiche, regione critica di livello α, eventuali condizioni di applicabilità.
2. Si valuti il p-value dei dati per il test introdotto e si tragga una conclusione.
3. La conclusione è forte o debole?

Si verifichi quindi l'affermazione di Pino.

4. Si imposti un opportuno test statistico, specificando variabili aleatorie di interesse, ipotesi statistiche, regione critica di livello α, eventuali condizioni di applicabilità.
5. Si valuti il p-value dei dati per il test introdotto e si tragga una conclusione.
6. La conclusione è forte o debole?

Quale macchinario conviene quindi ad Alessandro?

Risultati.

1.

$$X = \text{farina versata in un sacco dal macchinario di Nando}$$
$$Y = \text{farina versata in un sacco dal macchinario di Pino}$$
$$H_0 \ : \ \sigma_X \geq \sigma_Y \qquad H_1 \ : \ \sigma_X < \sigma_Y$$
$$R_\alpha \ : \ f_0 = \frac{s_X^2}{s_Y^2} < f_{1-\alpha,60,40} = \frac{1}{f_{\alpha,40,60}}$$

   Condizioni di applicabilità: le variabili devono avere distribuzione normale.

2.

$$f_0 = \frac{s_X^2}{s_Y^2} = 0.97 = \frac{1}{f_{\alpha,40,60}}$$
$$\frac{1}{f_0} = 1.03 = f_{\alpha,40,60} < 1.21 = f_{0.25,40,60}$$
$$\alpha > 0.25$$

   per cui il p-value è alto, maggiore del 25% (il valore esatto è pari a 45.53%), e i dati non forniscono una forte evidenza statistica a favore dell'ipotesi alternativa: non è provato che il macchinario di Nando sia più preciso del macchinario di Pino. Non possiamo quindi rifiutare l'ipotesi nulla: il macchinario di Nando non è più preciso del macchinario di Pino.

3. La conclusione è debole.

4.

$$X = \text{farina versata in un sacco dal macchinario di Nando}$$
$$Y = \text{farina versata in un sacco dal macchinario di Pino}$$
$$H_0 \ : \ \sigma_X = \sigma_Y \qquad H_1 \ : \ \sigma_X \neq \sigma_Y$$
$$R_\alpha \ : \ f_0 = \frac{s_X^2}{s_Y^2} < f_{1-\alpha/2,60,40} = \frac{1}{f_{\alpha/2,40,60}} \quad \text{oppure} \quad f_0 = \frac{s_X^2}{s_Y^2} > f_{\alpha/2,60,40}$$

   Condizioni di applicabilità: le variabili devono avere distribuzione normale.

5. Poichè $f_0 = 0.97 < 1$ si trova il p-value imponendo l'uguaglianza sulla prima condizione di R_α:

$$f_0 = \frac{s_X^2}{s_Y^2} = 0.97 = \frac{1}{f_{\alpha/2,40,60}}$$
$$\frac{1}{f_0} = 1.03 = f_{\alpha/2,40,60} < 1.21 = f_{0.25,40,60}$$
$$\alpha > 0.5$$

   per cui il p-value è alto, maggiore del 50% (il valore esatto è pari a 91.06%), e i dati non forniscono una forte evidenza statistica per rifiutare l'ipotesi nulla: i macchinari hanno la stessa precisione.

6. La conclusione è debole, ma rafforzata dall'alto p-value.

Conviene acquistare il macchinario di Pino: stessa precisione, ma più economico.

Problema 3. Quest'estate Paperino, avendo ben due mesi di vacanza, decide di fare un giro dell'Italia, memore dei racconti di suo cugino Gastone che l'anno precedente, di ritorno dallo stesso viaggio, aveva detto di non aver trovato un solo giorno di pioggia. Purtroppo Paperino non aveva fatto i conti con la terribile estate 2014 (e con la sua sfortuna) e, su 60 giorni di viaggio, è quasi sempre piovuto a dirotto, ad esclusione di pochi giorni di pioggia leggera. La seguente tabella riporta i mm di pioggia giornalieri caduti nei 60 giorni del viaggio di Paperino:

Classi	Frequenze assolute
$[0; 0.5)$	2
$[3.5; 4)$	3
$[4; 4.5)$	11
$[4.5; 5)$	4
$[5; 5.5)$	20
$[5.5; 6)$	18
$[6; 6.5)$	2

a) Rappresentare con un istogramma la distribuzione dei dati. Utilizzare le stesse classi della tabella.

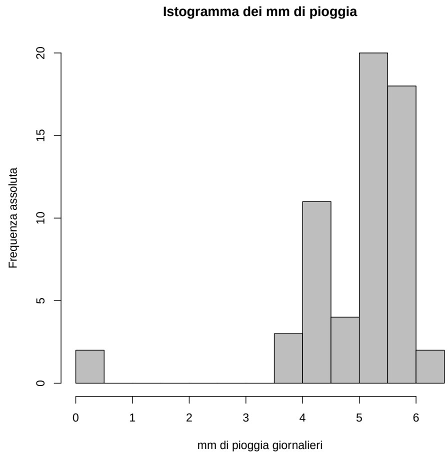
b) Indicare a quali classi appartengono il primo quartile Q_1 e il terzo quartile Q_3.

c) Quant'è il valore massimo che può assumere il range interquartile IQR? E il valore minimo?

d) Indicare il numero di outlier inferiori e il numero di outlier superiori.

e) Determinare dei valori approssimati di Q_1, Q_3 e IQR.

Risultati.



b) Data la tabella con le frequenze cumulate:

Classi	Frequenze assolute	Frequenze relative	Frequenze cumulate
$[0; 0.5)$	2	0.033	0.033
$[3.5; 4)$	3	0.05	0.083
$[4; 4.5)$	11	0.183	0.267
$[4.5; 5)$	4	0.067	0.333
$[5; 5.5)$	20	0.333	0.667
$[5.5; 6)$	18	0.3	0.967
$[6; 6.5)$	2	0.033	1

Tabella 1: Tabella delle frequenze per i mm di pioggia

si ricava che $Q_1 \in [4; 4.5)$ e $Q_3 \in [5.5; 6)$.

c) Il valore massimo dell'IQR è $6 - 4 = 2$, nel caso Q_1 e Q_3 coincidano con gli estremi esterni delle loro classi di appartenza, mentre il valore minimo è $5.5 - 4.5 = 1$, nel caso Q_1 e Q_3 coincidano con gli estremi interni.

d) Dato che il baffo inferiore può estendersi fino ad un estremo minimo $Q_1 - 1.5 \cdot IQR$ compreso tra $4 - 1.5 \cdot 2 = 1$ e $4.5 - 1.5 * 1 = 3$, sicuramente tutti e soli gli outlier inferiori sono i due dati appartenenti alla prima classe. Quindi abbiamo due outlier inferiori.

Dato che il baffo superiore può estendersi fino ad un estremo massimo $Q_3 + 1.5 \cdot IQR$ compreso tra $5.5 + 1.5 = 7$ e $6 + 3 = 9$, possiamo affermare che non ci sono outlier superiori.

e) $Q_1 \simeq 4 + (0.25 - 0.083) * \frac{0.5}{0.267 - 0.083} = 4.5 + 0.45 = 4.45$

$Q_3 \simeq 5.5 + (0.75 - 0.667) * \frac{0.5}{0.967 - 0.667} = 5.5 + 0.14 = 5.64$

$IQR \simeq 5.64 - 4.45 = 1.19.$