PROBABILITÀ E STATISTICA PER L'INGEGNERIA
2a Prova Intermedia
28 Maggio 2021

SOLUZIONI

DOMANDE E RISPOSTE

1. Si supponga che l'insieme di dati $\mathcal{D} = \{d_1, \ldots, d_{10}\}$ descriva il ritardo, in minuti, con cui un treno chiude le porte e parte, durante dieci mattine, rispetto all'orario previsto. Se i ritardi hanno una distribuzione normale $\mathcal{N}(\mu, \sigma^2)$, quanto valgono le stime più verosimili del ritardo medio $\hat{\mu}$ e della sua deviazione standard $\hat{\sigma}$?

   Risposta:

   - $\hat{\mu} = \frac{1}{10}\sum_{i=1}^{10} d_i$ 3.050
   - $\hat{\sigma} = \sqrt{\frac{1}{10}\sum_{i=1}^{10}(d_i - \hat{\mu})^2}$ 2.962

2. Si consideri la situazione del problema 1. Calcolare l'intervallo di confidenza al 95% del ritardo medio μ, supponendo che la varianza sia nota ($\sigma^2 = 4$).

   Risposta: 1.810 4.290
   La statistica $Z = \frac{\hat{\mu}-\mu}{\sigma/\sqrt{n}}$ è distribuita normalmente $\mathcal{N}(0,1)$, quindi: $\mu \in \left(\hat{\mu} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \hat{\mu} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$ con $n = 10$, $\sigma = 2$, $\alpha/2 = .025$ e $z_{.025} \approx 1.96$, come ricavabile dalla tabella per la distribuzione normale.

3. Si consideri la situazione del problema 1. Calcolare l'intervallo di confidenza al 95% del ritardo medio μ, supponendo che la varianza non sia nota.

   Risposta: S= 3.122
   In questo caso è possibile sostituire la varianza non nota σ^2 con la varianza campionaria $s^2 = \frac{1}{9}\sum_{i=1}^{10}(d_i - \hat{\mu})^2$. La statistica $T = \frac{\hat{\mu}-\mu}{s/\sqrt{n}}$ non è più normale ma segue una distribuzione t di Student con $n-1$ gradi di libertà. Di conseguenza, $\mu \in \left(\hat{\mu} - t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}, \hat{\mu} + t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}\right)$ con $n = 10$, $\alpha/2 = .025$ e $t_{.025,9} \approx 2.26$, come ricavabile dalla tabella per la distribuzione t di Student con 9. (0.817, 5.283)

4. Considerare i due intervalli di confidenza ottenuti negli esercizi precedenti. Sono diversi? Quale dei due intervalli ha ampiezza maggiore? Perchè?

   Risposta:
   In genere gli intervalli sono diversi, infatti si può dimostrare che l'ampiezza media dell'intervallo di confidenza è maggiore quando la varianza non è nota. Intuitivamente, ciò è dovuto al fatto che la stima della media diventa più "incerta" perché si ha a disposizione meno informazione (la varianza, appunto). Il fatto che l'ampiezza media di un intervallo sia maggiore dell'ampiezza media dell'altro non implica, ovviamente, che questo succeda sempre: per alcuni insiemi di dati può capitare esattamente l'opposto.

5. Sia H_0 l'ipotesi nulla in un test di ipotesi. Si supponga che il p-value sia maggiore del livello di significatività fissato ($p > \alpha$). Si può affermare che l'ipotesi H_0 è sicuramente vera?

   Risposta:
   No. Possiamo solo affermare che, per il livello di significatività fissato, i dati non confutano l'ipotesi. Cambiando il valore di α o con un diverso set di dati il risultato potrebbe cambiare.

6. Si consideri la situazione del problema 1. Verificare l'ipotesi che il treno abbia un ritardo medio maggiore o uguale a 5 minuti ($H_0 : \mu \geq \mu_0 = 5$) con livelli di significatività $\alpha_1 = 0.1$ e $\alpha_2 = 0.01$.

   Risposta: T = -1.98
   L'ipotesi può essere verificata supponendo che sia vera e verificando se la statistica $T = \frac{\hat{\mu}-\mu}{s/\sqrt{n}}$ restituisce un valore estremo (cioè poco probabile) rispetto al livello di significatività fissato. Nel nostro caso si suppone $\mu = 5$ e si sostituiscono i valori ricavati dai dati per n, s e $\hat{\mu}$ (esercizio 3). L'ipotesi è rifiutata se $T < -t_{\alpha,n-1}$ ovvero $T < -t_{0.1,9}$ o $T < -t_{0.01,9}$ nei due casi richiesti.

   -1.38 -2.821
   Reject ok

7. Si consideri la situazione del problema 1 ma si supponga che i ritardi siano distribuiti non più secondo una normale ma siano compresi tra 0 e θ con distribuzione uniforme $\mathcal{U}(0,\theta)$. Calcolare le stime a massima verosimiglianza di θ e del ritardo medio del treno.

   Risposta:
   $\hat{\theta} = \max(d_1,\ldots,d_{10})$ e il ritardo medio è $\hat{\theta}/2$ $\quad \hat{\vartheta} = 9.0 \quad \hat{\vartheta}/2 = 4.5$

8. Si consideri l'insieme di coppie di variabili (x_i, Y_i), con $i = 1,\ldots,10$, $x_i = i$ e Y_i come definiti precedentemente. Calcolare la stima della retta di regressione $Y(x) = A + Bx$ secondo il metodo dei minimi quadrati.

   Risposta:
   Per prima cosa si calcolino le seguenti quantità:
   $\overline{x} = 1/10\sum_{i=1}^{10} x_i$ $\quad 5.50$
   $\overline{Y} = 1/10\sum_{i=1}^{10} Y_i$ $\quad 3.05$
   $S_{xY} = \sum_{i=1}^{10} x_iY_i - n\overline{x}\,\overline{Y}$ $\quad 77.8$
   $S_{xx} = \sum_{i=1}^{10} x_i^2 - n\overline{x}^2$ $\quad 82.5$
   $S_{YY} = \sum_{i=1}^{10} Y_i^2 - n\overline{Y}^2$ $\quad 82.7$
   Quindi:
   $B = \frac{S_{xY}}{S_{xx}}$ $\quad 0.94$
   $A = \overline{Y} - B\overline{x}$ $\quad -2.133$

9. Si consideri la retta di regressione dell'esercizio precedente. Calcolare il valore $Y(x_0)$, con $x_0 = 11$. Supponendo che gli errori siano indipendenti e seguano una distribuzione normale a media nulla e varianza non nota, calcolare l'intervallo di predizione per $Y(x_0)$ con confidenza del 95%.

   Risposta:
   $Y(11) = A + B\cdot 11$ $\quad 8.233$
   Quindi si calcoli:
   $SS_R = \frac{S_{xx}S_{YY} - S_{xY}^2}{S_{xx}}$ $\quad 14.452$
   $(4.480, 11.987)$
   L'intervallo di predizione al 95% è dato da $Y(11) \pm \underbrace{t_{0.025,8}}_{2.306}\sqrt{\left[1 + \frac{1}{10} + \frac{(\overline{x}-11)^2}{S_{xx}}\right]\frac{SS_R}{8}}$

10. Spiegare il significato della *funzione di rischio* $\lambda(t) = \frac{f(t)}{1-F(t)}$.

    Risposta:
    La funzione di rischio $\lambda(t)$ rappresenta la densità condizionale di probabilità che un oggetto di età t si guasti "nel prossimo istante".

11. Sia $Y = \max(Y_1,\ldots,Y_{10})$. Si considerino due costruttori di automobili A e B e siano $\lambda_A = 1/Y$ e $\lambda_B = 2/Y$ i rispettivi tassi di guasto delle automobili prodotte. Supponendo tempi di vita esponenziali, calcolare i tempi di vita media, θ_A e θ_B, delle automobili dei due costruttori.

    Risposta:
    $\theta_A = Y$ e $\theta_B = Y/2$ $\quad \vartheta_A = 9 \quad \vartheta_B = 4.5$

12. Si consideri la situazione del caso precedente. Calcolare la probabilità che un'automobile del costruttore A, vecchia di 15 anni ma funzionante, continui a funzionare ancora per 5 anni. Calcolare l'analoga probabilità per un'automobile del costruttore B.

    Risposta:
    Se P è la probabilità che l'automobile vecchia di 15 anni arrivi a 20 anni, allora $P = e^{-\int_{15}^{20}\lambda(t)dt}$. Quindi $P_A = e^{-5/Y}$ e $P_B = e^{-10/Y}$, ovvero $P_B = P_A^2$.

    $P_A = e^{-5/9} = 0.57$

    $P_B = e^{-10/9} = P_A^2 = 0.33$