Politecnico di Milano - Scuola di Ingegneria Industriale

II Appello di Statistica per Ingegneria Energetica
5 settembre 2012

Cognome, Nome e Numero di matricola:

Problema 1. La Xair ha installato sui suoi aerei dei rilevatori ottici in grado di misurare la larghezza dei bagagli da stiva imbarcati. Gli addetti dell'aeroporto hanno costruito l'istogramma riportato in Figura 1, utilizzando i bagagli presenti sul volo Xair34 delle ore 9:00 diretto a New York.

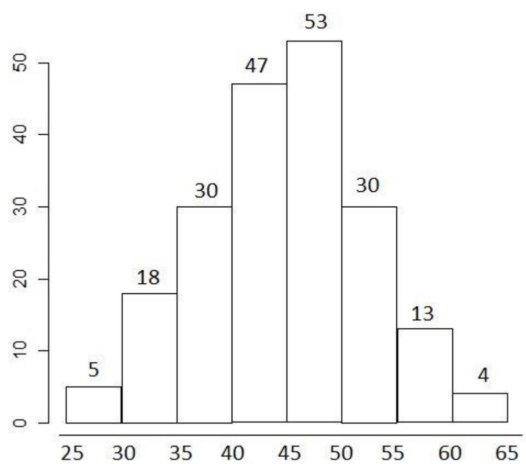

Figura 1: Istogramma delle larghezze dei bagagli presenti sul volo Xair34

(a) Costruire la tabella delle frequenze relative dei bagagli presenti sul volo Xair34 e individuare la classe modale.

(b) Calcolare approssimativamente la larghezza media dei bagagli presenti sul volo Xair34

Per ridurre il consumo di carburante sarebbe opportuno posizionare i bagagli di classe H (peso maggiore di 15 kg) in prossimità delle ali, mentre quelli di classe L (peso inferiore a 15 kg) in testa o coda. Per questioni di spazio nella stiva, la Xair richiede che almeno l'85% dei bagagli di classe H non superi i 57cm di larghezza e che almeno il 90% dei bagagli di classe L non superi i 47cm.

(c) Dagli istogrammi riportati in Figura 2, possiamo dire che i passeggeri del volo Xair34 hanno rispettato le direttive della compagnia?

Risultati.

(a) Contando le frequenze assolute scopro che ci sono $n = 200$ bagagli. Divido le frequenze assolute per 200 trovando le frequenze relative. La classe modale é [45;50)

Classe	[25;30)	[30;35)	[35;40)	[40;45)	[45;50)	[50;55)	[55;60)	[60;65)
f_i	0.025	0.090	0.150	0.235	0.265	0.150	0.065	0.020

Tabella 1: Tabella delle frequenze relative

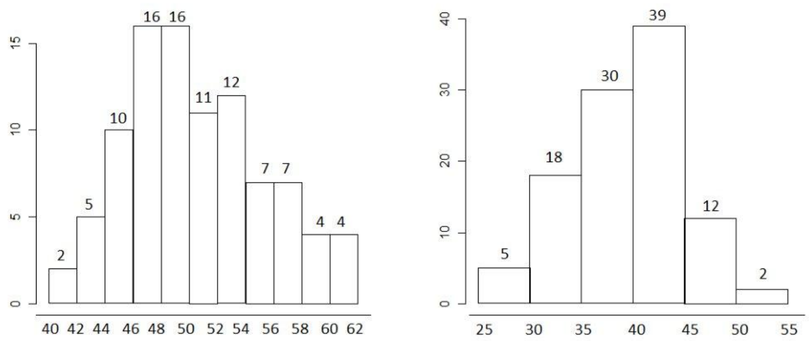

Figura 2: A sinistra: istogramma delle larghezze dei bagagli di classe H presenti sul volo Xair34. A destra: istogramma delle larghezze dei bagagli di classe L presenti sul volo Xair34.

(b) Per calcolare la media sommo i valori centrali delle classi pesandoli per le frequenze relative

$$\bar{x} \ = \ \sum_{i=1}^{8} m_i f_i \ = \ 27.5 \cdot 0.025 + 32.5 \cdot 0.090 + 37.5 \cdot 0.15 + 42.5 \cdot 0.235 +$$

$$47.5 \cdot 0.265 + 52.5 \cdot 0.15 + 57.5 \cdot 0.065 + 62.5 \cdot 0.02 \ = \ 44.675$$

La larghezza media é $\bar{x} = 44.675cm$

(c) Entrambe le direttive sono rispettate, infatti la frequenza dei bagagli di classe H che non superano i 57cm di larghezza è pari a

$$\sum_{i=1}^{8} f_i \ + \ \frac{57-56}{2} \cdot f_9 \ = \ 0.8404 + 0.5 \cdot 0.0745 \ = \ 0.8776 \ \geq \ 0.85$$

mentre quella dei bagagli di classe L che non superano i 47cm di larghezza è pari a

$$\sum_{i=1}^{4} f_i \ + \ \frac{47-45}{5} \cdot f_5 \ = \ 0.8679 + 0.4 \cdot 0.1132 \ = \ 0.9132 \ \geq \ 0.90$$

Problema 2. Il dott. Fantastic ha messo a punto un nuovo strumento in grado di misurare temperature molto elevate. Grazie alla sua intelligenza fuori dal comune, è riuscito a sapere che il suo strumento commette un errore casuale Δ la cui densità di probabilità ha la seguente forma:

$$f_\Delta(t) = ce^{-\lambda|t|}, \quad \lambda > 0.$$

Purtroppo il dott. Fantastic è un po' debole in probabilità, pertanto vi sottopone le seguenti richieste.

(a) Trovare l'espressione di c, in funzione di λ, tale per cui $f_\Delta(t)$ sia un densità di probabilità.

(b) Determinare il valore di λ per il quale la probabilità che Δ sia compreso tra -1 e 1 sia pari a 0.6321. Disegnare un grafico qualitativo di $f_\Delta(t)$.

Per il valore di λ trovato al punto (b), calcolare:

(c) la funzione di ripartizione $F_\Delta(t)$ e disegnarne un grafico qualitativo;

(d) media, mediana e varianza di Δ;

(e) la percentuale di misure effettuate dallo strumento che hanno un errore con valore assoluto inferiore a $2\sqrt{2}$;

(f) il quantile di ordine 25% degli errori dello strumento.

Soluzione.

(a) $\int_{-\infty}^{+\infty} f_\Delta(t)\mathrm{d}t = 1 \Rightarrow c = \frac{\lambda}{2}$.

(b) $0.6321 = P[-1 \leq \Delta \leq 1] = \int_{-1}^{1} f_\Delta(t) = 2\int_0^1 \frac{\lambda}{2}e^{-\lambda t} = 1 - e^{-\lambda} \Rightarrow \lambda \simeq 1$.

(c) Per $\lambda = 1$, si ha:

$$F_\Delta(t) = \int_{-\infty}^{t} f_\Delta(x)\mathrm{d}x = \begin{cases} [\frac{1}{2}e^x]_{-\infty}^t = \frac{1}{2}e^t & t \leq 0 \\ \frac{1}{2} + [-\frac{1}{2}e^{-x}]_0^t = 1 - \frac{1}{2}e^t & t > 0 \end{cases}$$

(d) $E[\Delta] = 0$ e mediana$(\Delta) = 0$ per simmetria. $Var(\Delta) = \int_{-\infty}^{+\infty} t^2 \frac{1}{2}e^{-|t|}\mathrm{d}t = 2\int_0^{+\infty} t^2 \frac{1}{2}e^{-t}\mathrm{d}t = 2$.

(e) $P[|\Delta| < 2\sqrt{2}] = P[-2\sqrt{2} \leq \Delta \leq 2\sqrt{2}] = 1 - 0.5e^{-2\sqrt{2}} - 0.5e^{-2\sqrt{2}} = 0.9409 \approx 94.09\%$.

(f) $F_\Delta(t) = 0.25 \Rightarrow 0.5e^t = 0.25 \Rightarrow t = \log(0.5) = -0.693$, quindi $q_{0.25} = -0.693$.

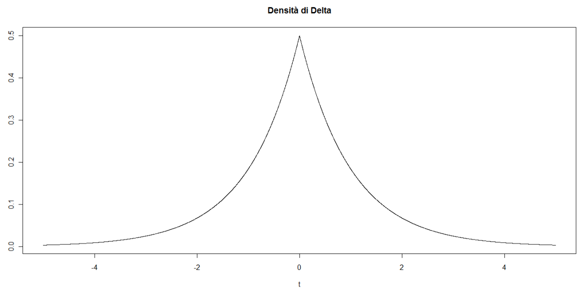

Figura 3: Grafico della funzione densità $f_\Delta(t)$

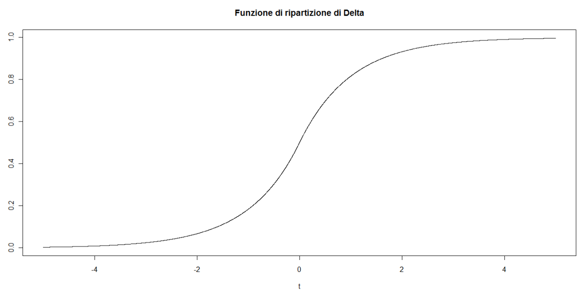

Figura 4: Grafico della funzione di ripartizione $F_\Delta(t)$

Problema 3. K. P. Bioteng è un famoso ingegnere biomedico curioso di testare se l'effetto della tachipirina sia lo stesso sugli uomini e sulle donne. In particolare vuole dimostrare che l'effetto è migliore sulle donne, al contrario di quanto sempre affermato dall'ingegner Bionucci, che ha convinto la comunità scientifica del fatto che l'effetto della tachipirina non dipende dal sesso dell'individuo. A tal scopo raccoglie un campione di 16 donne e 13 uomini e, per ognuno dei pazienti considerati, registra la diminuzione della temperatura corporea X (in °C) a tre ore di distanza dall'assunzione di una compressa di tachipirina da 500 mg. Le analisi effettuate hanno fatto registrare i seguenti valori di medie e varianze campionarie:

$$\overline{x}_D = 1.12 \qquad \overline{x}_U = 1.02 \qquad s_D^2 = 0.0069 \qquad s_U^2 = 0.0063$$

Assumiamo che le osservazioni siano indipendenti e che $X_D \sim N(\mu_D, \sigma_D^2)$ e $X_U \sim N(\mu_U, \sigma_U^2)$, dove X_D e X_U sono variabili aleatorie che rappresentano la diminuzione della temperatura corporea per gli uomini e per le donne a seguito dell'assunzione di una compressa di tachipirina da 500 mg.

a) Le varianze di X_D e X_U possono essere considerate uguali? Per rispondere eseguire un opportuno test con livello di significatività del 5%, specificando ipotesi nulla, ipotesi alternativa, regione critica e conclusione.

b) L'ingegner Bioteng è in grado di smentire la tesi dell'ingegner Bionucci? Per rispondere eseguire un opportuno test, specificando ipotesi nulla, ipotesi alternativa, regione critica, p-value e conclusione.

Si supponga ora che le deviazioni standard siano note e diverse, ovvero $\sigma_D = 0.08$ e $\sigma_U = 0.07$. Si supponga inoltre di dover analizzare i risultati di una nuova indagine condotta su un campione con un numero di uomini doppio del numero di donne.

c) Quanto deve essere il numero di donne considerate nell'esperimento affinché l'intervallo di confidenza bilatero di livello 90% per la differenza $\mu_D - \mu_U$ abbia una lunghezza minore di 0.05?

Risultati.

a) $H_0 : \sigma_D^2 = \sigma_U^2$ vs. $H_1 : \sigma_D^2 \neq \sigma_U^2$
$F_0 = \frac{S_D^2}{S_U^2} \sim F_{n_D-1,n_U-1} \to \text{RC} = \{f_0 < f_{\frac{\alpha}{2},n_D-1,n_U-1}\} \cup \{f_0 > f_{1-\frac{\alpha}{2},n_D-1,n_U-1}\}$.
Con i dati ottenuti ricaviamo $f_0 = \frac{s_D^2}{s_U^2} = 1.095$. Sapendo che, dalle tavole, $f_{0.025,15,12} = \frac{1}{f_{0.975,12,15}} = \frac{1}{2.96} = 0.338$ e $f_{0.975,15,12} = 3.18$, non possiamo rifiutare l'ipotesi nulla ad un livello di significatività del 5%.

b) $H_0 : \mu_D - \mu_U = 0$ vs. $H_1 : \mu_D - \mu_U > 0$
$T_0 = \frac{\overline{X}_D - \overline{X}_U}{S_p\sqrt{\frac{1}{n_D}+\frac{1}{n_U}}} \sim T_{n_D+n_U-2} \to \text{RC} = \{t_0 > t_{1-\alpha,n_D+n_U-2}\}$ dove $S_p^2 = \frac{(n_D-1)S_D^2+(n_U-1)S_U^2}{n_D+n_U-2}$.
Con i dati ottenuti ricaviamo $S_p^2 = 0.0066 \to t_0 = 3.3$. Poichè dalle tavole ricaviamo $3.057 = t_{0.9975,27} < 3.3 < t_{0.999,27} = 3.421$, possiamo dire che $0.001 < p-value < 0.0025$. Quindi abbiamo forte evidenza per rifiutare H_0 e di conseguenza possiamo dire che l'ingegner Bioteng è riuscito a dimostrare la sua congettura smentendo l'ingegner Bionucci.

c) Nel caso in cui $n_U = 2n_D$, $IC_{0.9}(\mu_D - \mu_U) = \left[\overline{X}_D - \overline{X}_U \pm z_{0.95}\sqrt{\frac{\sigma_D^2}{n_D} + \frac{\sigma_U^2}{2n_D}}\right]$.
Quindi $L = 2z_{0.95}\sqrt{\frac{\sigma_D^2}{n_D} + \frac{\sigma_U^2}{2n_D}} \leq 0.05 \to n_D \geq 38.31 \to n_D \geq 39$.