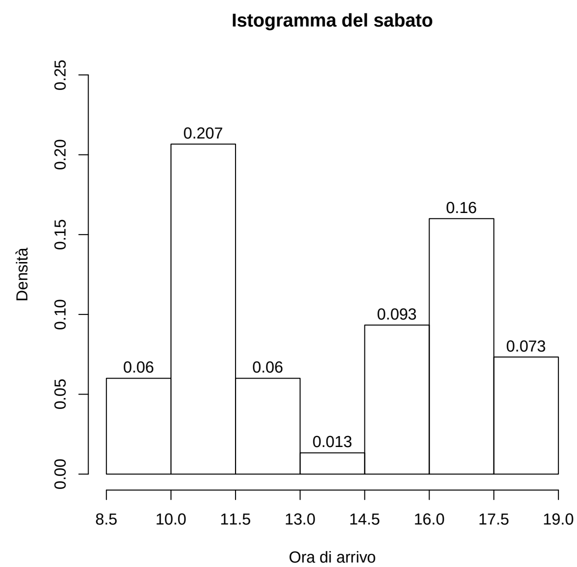
**Politecnico di Milano - Scuola di Ingegneria Industriale e dell'Informazione**

I Prova in Itinere di Statistica per Ingegneria Energetica
26 novembre 2015



*Cognome, Nome e Numero di matricola:*

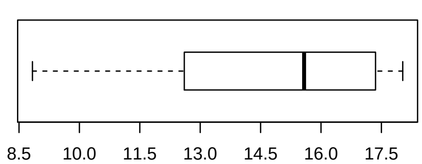
**Problema 1**. Alberto ha deciso di andare a tagliarsi i capelli da *Il Milanese Parrucchieri Uomo* e, dato che durante la settimana è impegnato al lavoro fino a tardi, guarda gli orari di apertura del sabato. Dello stesso giorno scarica dal sito web dell'attività l'ora di arrivo dei clienti per decidere quando andare. Il seguente istogramma riassume la distribuzione dei dati raccolti da Alberto:



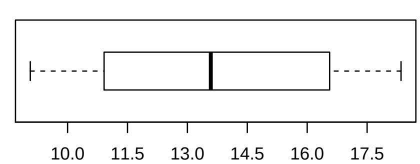

(a) Costruire la tabella delle frequenze relative e cumulate dell'ora di arrivo dei clienti.

(b) Stabilire a quale classe appartengono la mediana, il primo e il terzo quartile.

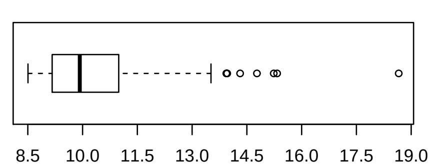
(c) Dei seguenti boxplot, uno è proprio quello dei dati raccolti. Quale? Perché?



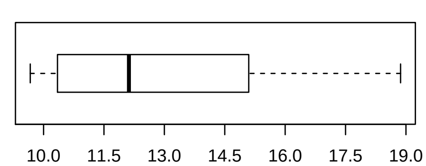

(a)




(b)




(c)




(d)

Alberto scopre di avere un impegno il sabato e decide di prendere mezza giornata di ferie il mercoledì successivo; di seguito i dati raccolti per tale giorno:

| | | | | |
|---|---|---|---|---|
| 8.652 | 10.748 | 14.078 | 15.459 | 15.539 |
| 15.739 | 15.797 | 15.798 | 16.041 | 16.080 |
| 16.261 | 16.665 | 16.732 | 16.780 | 16.876 |
| 16.989 | 17.250 | 17.459 | 17.459 | 17.675 |
| 17.935 | 18.176 | 18.331 | 18.440 | 18.827 |

(d) Calcolare mediana, quartili, range interquartile, outlier e disegnare il boxplot.

(e) Completare la tabella delle frequenze:

| Classi | Freq Ass | Freq Rel | Densità |
|---|---|---|---|
| [8.5,10) | | | |
| [10,11.5) | | | |
| [11.5,13) | | | |
| [13,14.5) | | | |
| [14.5,16) | | | |
| [16,17.5) | | | |
| [17.5,19) | | | |

(f) Costruire l'istogramma per la distribuzione di densità dell'ora di arrivo dei clienti il mercoledì.

(g) Evidenziare le differenze più evidenti tra la distribuzione del sabato e quella del mercoledì.

(h) Supponendo grossomodo stabili le distribuzioni, consigliereste ad Alberto di prendere le ferie il mercoledì mattina o il mercoledì pomeriggio?

**Risultati.**

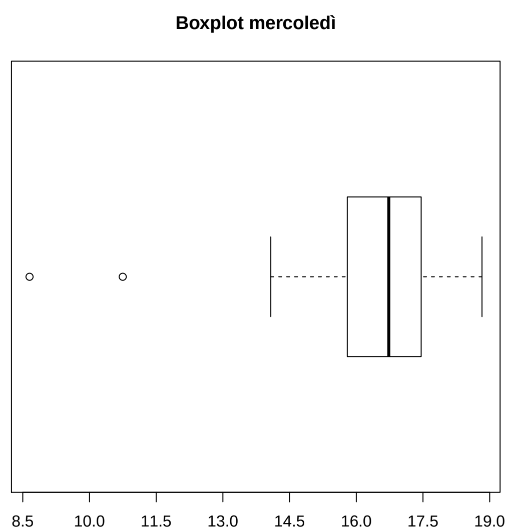
(a) La tabella delle frequenze del sabato è:

| Classi | Densità | Freq Rel | Freq Cum |
|---|---|---|---|
| [8.5,10) | 0.060 | 0.09 | 0.09 |
| [10,11.5) | 0.207 | 0.31 | 0.40 |
| [11.5,13) | 0.060 | 0.09 | 0.49 |
| [13,14.5) | 0.013 | 0.02 | 0.51 |
| [14.5,16) | 0.093 | 0.14 | 0.65 |
| [16,17.5) | 0.160 | 0.24 | 0.89 |
| [17.5,19) | 0.073 | 0.11 | 1 |

(b) La classe di appartenenza del quantile di ordine $p$ è la prima per cui si ha che la frequenza cumulata è maggiore o uguale a $p$. Abbiamo che $m \in [13, 14.5)$, $Q_1 \in [10, 11.5)$ mentre $Q_3 \in [16, 17.5)$.

(c) In base alle informazioni ottenute al punto (b) si ha che l'unico boxplot coerente coi risultati è il (b) dato che negli altri casi i quartili e/o la mediana cadono nelle classi sbagliate.

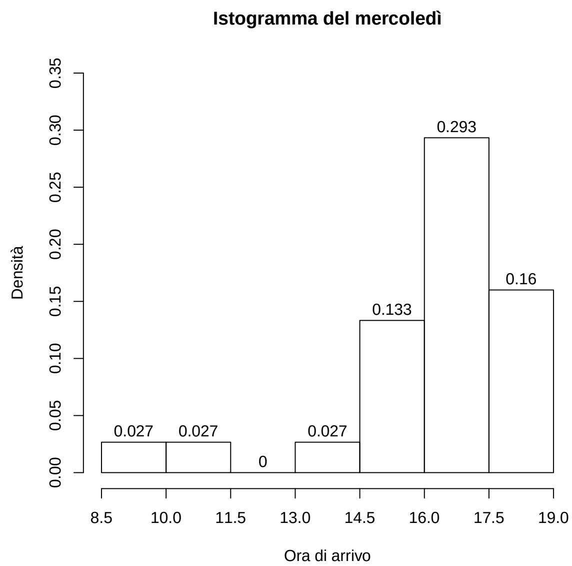
(d) $m = Q_2 = 16.732,\quad Q_1 = 15.797,\quad Q_2 = 17.459,\quad IQR = 1.662,\quad \text{outlier} = 8.652,\ 10.748.$




(e) La tabella delle frequenze del mercoledì è:

| Classi | Freq Ass | Freq Rel | Densità |
|---|---|---|---|
| [8.5,10) | 1 | 0.04 | 0.027 |
| [10,11.5) | 1 | 0.04 | 0.027 |
| [11.5,13) | 0 | 0 | 0 |
| [13,14.5) | 1 | 0.04 | 0.027 |
| [14.5,16) | 5 | 0.20 | 0.133 |
| [16,17.5) | 11 | 0.44 | 0.293 |
| [17.5,19) | 6 | 0.24 | 0.160 |

(f) L'istogramma per il mercoledì corrispondente alla tabella completata è:




Si potrebbe valutare di allargare le classi per ridurne il numero.

(g) La distribuzione del sabato è più simmetrica, priva di outlier, ma bimodale, corrispondente a una equiripartizione degli arrivi fra mattino e pomeriggio. La distribuzione del mercoledì è fortemente asimmetrica, con coda a sinistra e due outlier inferiori, corrispondente ad un affollamento degli arrivi nel pomeriggio con rare eccezioni la mattina.

(h) Consiglieremmo ad Alberto di prendere la mattina di ferie, visto che dal parrucchiere non ci va nessuno o quasi...

**Problema 2**. Eugenio e Simona si frequentano da tanti anni e Simona non è mai arrivata puntuale ad un appuntamento. In questo tempo Eugenio ha capito che il ritardo $X$ di Simona è casuale e che, misurato in minuti, ha distribuzione della forma

$$f(x) = \begin{cases} \dfrac{c}{x^4}, & \text{se } x \geq 1, \\ 0, & \text{altrimenti.} \end{cases}$$

(a) Calcolare $c$ tale per cui $f$ sia effettivamente una densità di probabilità. Tracciare quindi un grafico qualitativo di $f$.

(b) Calcolare la funzione di ripartizione di $X$ e tracciarne un grafico qualitativo.

(c) Quante volte, in percentuale, Simona ritarda più di 3 minuti?

(d) Di quanto può ritardare Eugenio, senza rischiare mai di arrivare dopo Simona?

(e) Quanto valgono media e varianza del ritardo di Simona?

(f) Eugenio e Simona si vedranno tutte le sere del mese di aprile 2016. Con quale probabilità in questo periodo Simona accumulerà un ritardo superiore ai 51 minuti?

**Risultati.**

(a) Affinché $f$ sia una densità deve essere:

$$\begin{cases} f(x) \geq 0 \quad \forall x \in \mathbb{R} \\ \int_{\mathbb{R}} f(x)dx = 1 \end{cases}.$$

Per quanto riguarda la prima condizione otteniamo $c \geq 0$. Per quanto riguarda la seconda abbiamo:

$$1 = \int_{\mathbb{R}} f(x)dx = c \int_1^{\infty} \frac{1}{x^4} dx = \frac{c}{3} \quad \Rightarrow \quad c = 3.$$

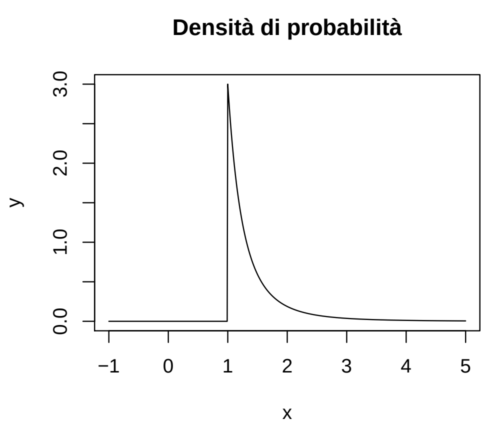
(b) La funzione di ripartizione di $X$ è

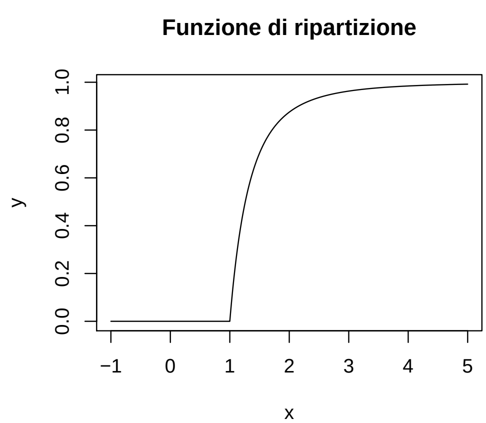
$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

Se $x < 1$ allora $f(x) = 0$ e quindi $F(x) = 0$. Se $x \geq 1$ abbiamo:

$$F(x) = 3 \int_1^x \frac{1}{t^4} dt = 1 - \frac{1}{x^3}.$$

I grafici di densità e fdr sono:




(c) $P(X > 3) = 1 - F(3) = \frac{1}{27} \approx 0.037 = 3.7\%$.

(d) Eugenio può ritardare di 1 minuto senza problemi, dato che $P(X > 1) = 1$.

(e) $\mathbb{E}[X] = 1.5$ e $\mathrm{Var}(X) = 0.75$.

(f) Considerando i ritardi di ogni appuntamento indipendenti tra loro, per il Teorema centrale del limite abbiamo:

$$Z = \frac{\sum_{i=1}^{30} X_i - 30 \cdot 1.5}{\sqrt{30} \cdot 0.866} \sim \mathcal{N}(0,1).$$

$$P\left(\sum_{i=1}^{30} X_i > 51\right) = 1 - P\left(\sum_{i=1}^{30} X_i \leq 51\right) = 1 - P\left(Z \leq \frac{51 - 45}{4.743}\right) =$$

$$= 1 - P(Z \leq 1.265) = 0.104$$

**Problema 3**. Al Dottor Baltar serve il valore di un certo angolo $\theta$. Dispone di uno strumento che misura gli angoli, in radianti, affetto da errore casuale $\epsilon$ gaussiano, centrato, di deviazione standard $\sigma_0 = 0.09$. Sia quindi $X$ una misura di $\theta$ effettuata con tale strumento.

(a) Scrivere $X$ in funzione di $\theta$ ed $\epsilon$, deducendo così la distribuzione di $X$.

Per ridurre l'effetto dell'errore di misura, il Dottor Baltar intende stimare $\theta$ sulla base di $n$ misure indipendenti dell'angolo.

(b) Introdurre un'opportuno stimatore $\widehat{\Theta}$ di $\theta$, dipendente da $n$ misure indipendenti $X_1, \ldots, X_n$, e specificarne la distribuzione.

(c) Quante misure servono a Baltar se vuole che l'errore standard di $\widehat{\Theta}$ sia inferiore a 0.05?

Al Dottor Baltar serve ora il valore di $c = \cos\theta$ e intende stimarlo con $\widehat{C} = \cos\widehat{\Theta}$.

(d) Determinare in maniera approssimata la distorsione e l'errore standard dello stimatore $\widehat{C}$.

Eseguendo 7 misure indipendenti di $\theta$, il Dottor Baltar trova i seguenti risultati

$$2.47, \quad 2.57, \quad 2.52, \quad 2.53, \quad 2.38, \quad 2.44, \quad 2.48.$$

(e) Sia per $\theta$ sia per $c$, determinare stima puntuale e relativo errore standard.

**Risultati.**

(a) $X = \theta + \epsilon \sim N(\theta, \sigma_0^2)$.

(b) $\widehat{\Theta} = \overline{X}_n \sim N\left(\theta, \dfrac{\sigma_0^2}{n}\right)$.

(c) $\dfrac{\sigma_0}{\sqrt{n}} < 0.05 \iff n > \dfrac{\sigma_0^2}{0.05^2} = 3.2 \iff n \geq 4$.

(d) Grazie al metodo delta (formula di propagazione degli errori):

distorsione $= \mathbb{E}[\widehat{C}] - c = \mathbb{E}[\cos\widehat{\Theta}] - c \simeq \cos\mathbb{E}[\widehat{\Theta}] - c = \cos\theta - \cos\theta = 0$,

$\mathrm{se}(\widehat{C}) = \sqrt{\mathrm{Var}(\widehat{C})} \simeq \dfrac{|\sin(\theta)|}{\sqrt{n}}\,\sigma_0$.

(e) $\widehat{\theta} = \overline{x}_7 = 2.48, \qquad \mathrm{se}(\widehat{\Theta}) = \dfrac{\sigma_0}{\sqrt{n}} = \dfrac{0.09}{\sqrt{7}} = 0.04,$

$\widehat{c} = \cos(\overline{x}_7) = -0.79, \qquad \mathrm{se}(\widehat{C}) \simeq \dfrac{|\sin(\widehat{\theta})|}{\sqrt{n}}\,\sigma_0 = 0.03.$