Statistica - 2ª lezione

23 febbraio 2023

Proprietà di s^2 e di s

- $[s^2] = [x_i]^2$, mentre $[s] = [\bar{x}] = [x_i]$.
- Vale la disuguaglianza di Chebyshev

$$\#\{i \in \{1, 2, \ldots, n\} \mid |x_i - \bar{x}| \geq k\, s\} \leq \frac{n-1}{k^2} \qquad \forall k > 0$$

- Vale la formula alternativa

$$s^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} x_i^2 - n \cdot \bar{x}^2 \right)$$

Proprietà di s^2 e di s

- $[s^2] = [x_i]^2$, mentre $[s] = [\bar{x}] = [x_i]$.
- Vale la disuguaglianza di Chebyshev

$$\#\{i \in \{1, 2, \ldots, n\} \mid |x_i - \bar{x}| \geq k\, s\} \leq \frac{n-1}{k^2} \qquad \forall k > 0$$

- Vale la formula alternativa

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n \cdot \bar{x}^2\right)$$

- Per dati discreti:

$$s_n^2 = \frac{n}{n-1} \sum_{\text{classi } k} (k - \overline{x})^2\, \mathrm{FR}(k)$$

Proprietà di s^2 e di s

- Per una trasformazione affine $y_i = ax_i + b$:

  $s_y^2 = a^2 s_x^2$

  $s_y = |a|\, s_x$

Proprietà di s^2 e di s

- Per una trasformazione affine $y_i = ax_i + b$:

  $s_y^2 = a^2 s_x^2$

  $s_y = |a|\, s_x$

  Infatti:

  $$s_y^2 = \frac{1}{n-1} \sum (y_i - \bar{y})^2$$

Proprietà di s^2 e di s

- Per una trasformazione affine $y_i = ax_i + b$:

  $s_y^2 = a^2 s_x^2$

  $s_y = |a|\, s_x$

  Infatti:

  $$s_y^2 = \frac{1}{n-1}\sum (y_i - \bar{y})^2$$

  $$= \frac{1}{n-1}\sum [(ax_i + b) - (a\bar{x} + b)]^2$$

Proprietà di s^2 e di s

- Per una trasformazione affine $y_i = ax_i + b$:

  $s_y^2 = a^2 s_x^2$

  $s_y = |a|\, s_x$

  Infatti:

  $$s_y^2 = \frac{1}{n-1} \sum (y_i - \bar{y})^2$$
  $$= \frac{1}{n-1} \sum [(ax_i + b) - (a\bar{x} + b)]^2$$
  $$= \frac{1}{n-1} \sum [a(x_i - \bar{x})]^2$$

Proprietà di s^2 e di s

- Per una trasformazione affine $y_i = ax_i + b$:

  $s_y^2 = a^2 s_x^2$

  $s_y = |a|\, s_x$

  Infatti:

$$s_y^2 = \frac{1}{n-1} \sum (y_i - \bar{y})^2$$

$$= \frac{1}{n-1} \sum [(ax_i + b) - (a\bar{x} + b)]^2$$

$$= \frac{1}{n-1} \sum [a(x_i - \bar{x})]^2 = \frac{a^2}{n-1} \sum (x_i - \bar{x})^2$$

Proprietà di s^2 e di s

- Per una trasformazione affine $y_i = ax_i + b$:

  $s_y^2 = a^2 s_x^2$

  $s_y = |a|\, s_x$

  Infatti:

$$
\begin{aligned}
s_y^2 &= \frac{1}{n-1}\sum (y_i - \bar{y})^2 \\
&= \frac{1}{n-1}\sum [(ax_i + b) - (a\bar{x} + b)]^2 \\
&= \frac{1}{n-1}\sum [a(x_i - \bar{x})]^2 = \frac{a^2}{n-1}\sum (x_i - \bar{x})^2 \\
&= a^2 s_x^2
\end{aligned}
$$

Proprietà di s^2 e di s

- Per una trasformazione affine $y_i = ax_i + b$:

$s_y^2 = a^2 s_x^2$

$s_y = |a|\, s_x$

Infatti:

$$
\begin{aligned}
s_y^2 &= \frac{1}{n-1}\sum (y_i - \bar{y})^2 \\
&= \frac{1}{n-1}\sum [(ax_i + b) - (a\bar{x} + b)]^2 \\
&= \frac{1}{n-1}\sum [a(x_i - \bar{x})]^2 = \frac{a^2}{n-1}\sum (x_i - \bar{x})^2 \\
&= a^2 s_x^2
\end{aligned}
$$

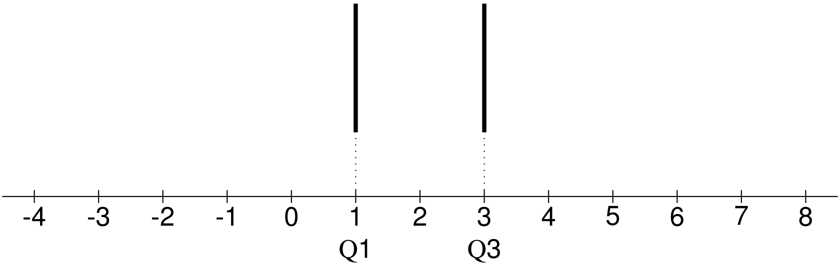
$$s_y = \sqrt{s_y^2} = \sqrt{a^2 s_x^2} = |a|\sqrt{s_x^2} = |a|\, s_x$$

Boxplot

Si disegna una riga (orizzontale o verticale):

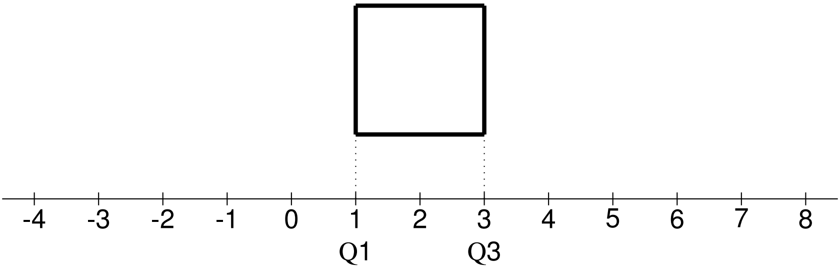
Boxplot

Si tirano due righe in corrispondenza di $\mathrm{Q1}$ e di $\mathrm{Q3}$:



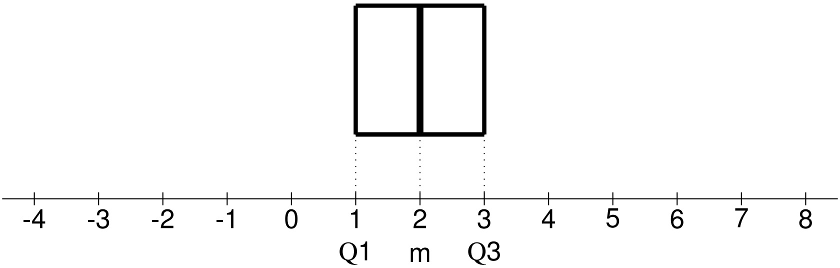
Boxplot

Si chiude la scatola:



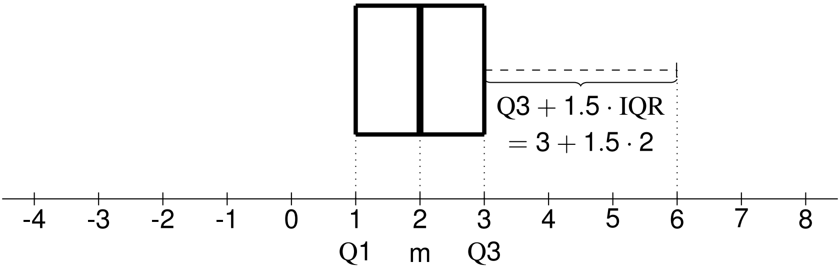
Boxplot

Si tira una righa spessa in corrispondenza di *m*:



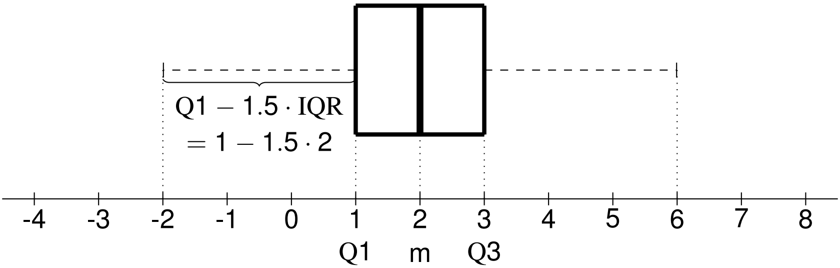
Boxplot

Si disegna un baffo da $\mathrm{Q3}$ a $\mathrm{Q3} + 1.5 \cdot \mathrm{IQR}$ ...



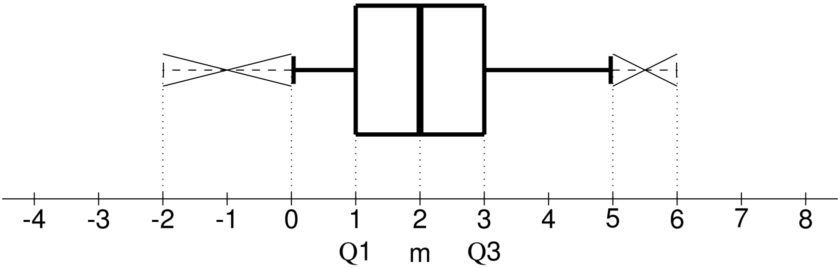
Boxplot

…e un altro baffo da $\mathrm{Q1}$ a $\mathrm{Q1} - 1.5 \cdot \mathrm{IQR}$:



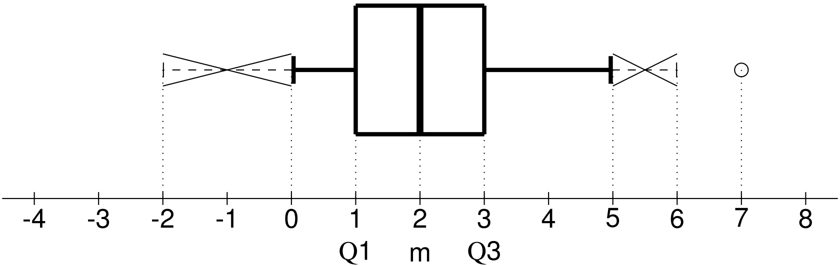
Boxplot

Dopodiché si tagliano i baffi in corrispondenza dei dati più esterni compresi entro di essi:



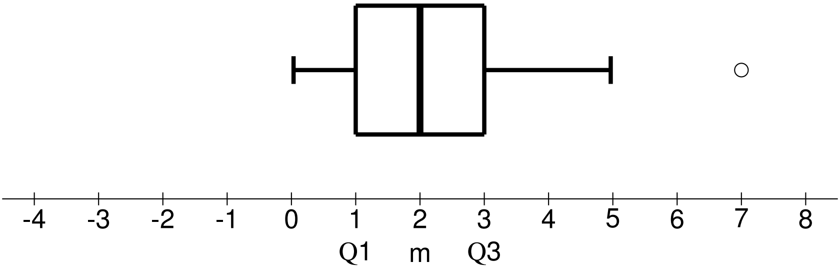
Boxplot

Eventuali dati esterni ai baffi si segnalano come *outlier*:

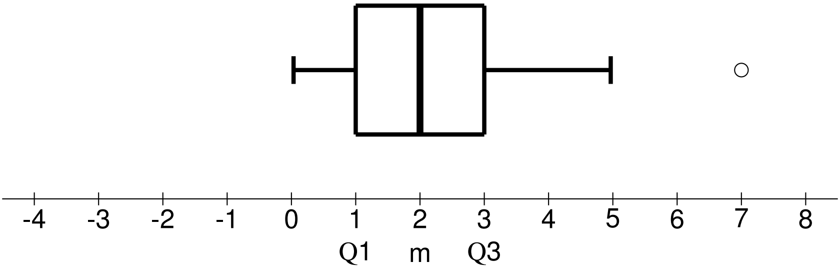


Boxplot

Ecco il risultato finale:



Boxplot



- Almeno il 50% dei dati è nella scatola
- Si vede un accenno di coda sulla destra

Dati continui

ESEMPIO: Abbiamo misurato le temperature massime giornaliere (in °C) dell'ultimo mese, ottenendo questi dati $x_1, x_2, \ldots, x_{30}$:

14.7	12.3	18.3	10.2	11.5	15.1	14.2	14.7	13.8	17.3
25.3	26.4	31.0	19.4	17.5	17.6	16.8	18.0	13.8	10.7
12.6	14.5	17.8	19.6	17.2	13.1	13.9	14.2	13.7	18.1

I valori possibili sono $S = \mathbb{R}$

Tutti gli indici di posizione e di dispersione si calcolano come prima:

$$\bar{x} = \frac{14.7 + 12.3 + \ldots + 18.1}{30} = 16.44\overline{3}$$

Dati continui

ESEMPIO: Abbiamo misurato le temperature massime giornaliere (in $^\circ\mathrm{C}$) dell'ultimo mese, ottenendo questi dati ordinati $x_{(1)}, x_{(2)}, \ldots, x_{(30)}$:

10.2	10.7	11.5	12.3	12.6	13.1	13.7	13.8	13.8	13.9
14.2	14.2	14.5	14.7	14.7	15.1	16.8	17.2	17.3	17.5
17.6	17.8	18.0	18.1	18.3	19.4	19.6	25.3	26.4	31.0

I valori possibili sono $S = \mathbb{R}$

Tutti gli indici di posizione e di dispersione si calcolano come prima:

$$\bar{x} = \frac{14.7 + 12.3 + \ldots + 18.1}{30} = 16.44\bar{3}$$

$$q_{0.5} = \frac{1}{2}\left(x_{(n\gamma)} + x_{(n\gamma+1)}\right) \qquad n\gamma = 30 \cdot 0.5 = 15 \in \mathbb{N}$$

$$= \frac{1}{2}\left(x_{(15)} + x_{(16)}\right) = \frac{1}{2}(14.7 + 15.1)$$

$$= 14.9$$

Dati continui

ESEMPIO: Abbiamo misurato le temperature massime giornaliere (in °C) dell'ultimo mese, ottenendo questi dati ordinati $x_{(1)}, x_{(2)}, \ldots, x_{(30)}$:

10.2	10.7	11.5	12.3	12.6	13.1	13.7	13.8	13.8	13.9
14.2	14.2	14.5	14.7	14.7	15.1	16.8	17.2	17.3	17.5
17.6	17.8	18.0	18.1	18.3	19.4	19.6	25.3	26.4	31.0

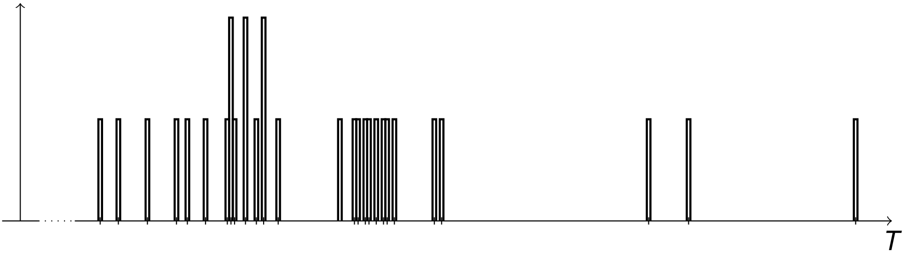
Ma come si disegna l'istogramma?

Dati continui

ESEMPIO: Abbiamo misurato le temperature massime giornaliere (in °C) dell'ultimo mese, ottenendo questi dati ordinati $x_{(1)}, x_{(2)}, \ldots, x_{(30)}$:

10.2	10.7	11.5	12.3	12.6	13.1	13.7	13.8	13.8	13.9
14.2	14.2	14.5	14.7	14.7	15.1	16.8	17.2	17.3	17.5
17.6	17.8	18.0	18.1	18.3	19.4	19.6	25.3	26.4	31.0

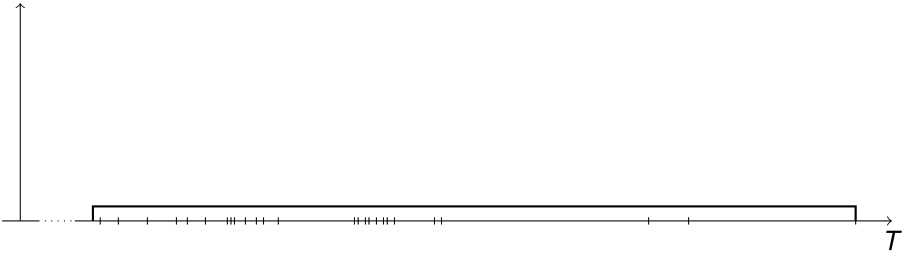
Se uso classi troppo numerose:



ESEMPIO: Abbiamo misurato le temperature massime giornaliere (in °C) dell'ultimo mese, ottenendo questi dati ordinati $x_{(1)}, x_{(2)}, \ldots, x_{(30)}$:

10.2	10.7	11.5	12.3	12.6	13.1	13.7	13.8	13.8	13.9
14.2	14.2	14.5	14.7	14.7	15.1	16.8	17.2	17.3	17.5
17.6	17.8	18.0	18.1	18.3	19.4	19.6	25.3	26.4	31.0

Se uso classi troppo poco numerose:



Dati continui

ESEMPIO: Abbiamo misurato le temperature massime giornaliere (in $^\circ$C) dell'ultimo mese, ottenendo questi dati ordinati $x_{(1)}, x_{(2)}, \ldots, x_{(30)}$:

10.2	10.7	11.5	12.3	12.6	13.1	13.7	13.8	13.8	13.9
14.2	14.2	14.5	14.7	14.7	15.1	16.8	17.2	17.3	17.5
17.6	17.8	18.0	18.1	18.3	19.4	19.6	25.3	26.4	31.0

Il numero giusto di classi è

$$\#\text{classi} = \sqrt{n}$$

oppure

$$\#\text{classi} = 1 + \log_2 n = 1 + \frac{\log_z n}{\log_z 2}$$

Dati continui

ESEMPIO: Abbiamo misurato le temperature massime giornaliere (in °C) dell'ultimo mese, ottenendo questi dati ordinati $x_{(1)}, x_{(2)}, \ldots, x_{(30)}$:

10.2	10.7	11.5	12.3	12.6	13.1	13.7	13.8	13.8	13.9
14.2	14.2	14.5	14.7	14.7	15.1	16.8	17.2	17.3	17.5
17.6	17.8	18.0	18.1	18.3	19.4	19.6	25.3	26.4	31.0

Il numero giusto di classi è

$$\#\text{classi} = \sqrt{n} = \sqrt{30} \simeq 5.477$$

oppure

$$\#\text{classi} = 1 + \log_2 n = 1 + \frac{\log_z n}{\log_z 2} = 1 + \frac{\ln 30}{\ln 2} \simeq 5.907$$

$$\Rightarrow \quad \text{scelgo 6 classi di ampiezza } \frac{31.0 - 10.2}{6} = 3.5\overline{6} \simeq 4$$

ESEMPIO: Abbiamo misurato le temperature massime giornaliere (in °C) dell'ultimo mese, ottenendo questi dati ordinati $x_{(1)}, x_{(2)}, \ldots, x_{(30)}$:

10.2	10.7	11.5	12.3	12.6	13.1	13.7	13.8	13.8	13.9
14.2	14.2	14.5	14.7	14.7	15.1	16.8	17.2	17.3	17.5
17.6	17.8	18.0	18.1	18.3	19.4	19.6	25.3	26.4	31.0

Classi			
$(10, 14]$			
$(14, 18]$			
$(18, 22]$			
$(22, 26]$			
$(26, 30]$			
$(30, 34]$			

ESEMPIO: Abbiamo misurato le temperature massime giornaliere (in °C) dell'ultimo mese, ottenendo questi dati ordinati $x_{(1)}, x_{(2)}, \ldots, x_{(30)}$:

10.2	10.7	11.5	12.3	12.6	13.1	13.7	13.8	13.8	13.9
14.2	14.2	14.5	14.7	14.7	15.1	16.8	17.2	17.3	17.5
17.6	17.8	18.0	18.1	18.3	19.4	19.6	25.3	26.4	31.0

Classi	FA		
$(10, 14]$	10		
$(14, 18]$			
$(18, 22]$			
$(22, 26]$			
$(26, 30]$			
$(30, 34]$			

Dati continui

ESEMPIO: Abbiamo misurato le temperature massime giornaliere (in °C) dell'ultimo mese, ottenendo questi dati ordinati $x_{(1)}, x_{(2)}, \ldots, x_{(30)}$:

10.2	10.7	11.5	12.3	12.6	13.1	13.7	13.8	13.8	13.9
14.2	14.2	14.5	14.7	14.7	15.1	16.8	17.2	17.3	17.5
17.6	17.8	18.0	18.1	18.3	19.4	19.6	25.3	26.4	31.0

Classi	FA		
$(10, 14]$	10		
$(14, 18]$	13		
$(18, 22]$			
$(22, 26]$			
$(26, 30]$			
$(30, 34]$			

Dati continui

ESEMPIO: Abbiamo misurato le temperature massime giornaliere (in $^{\circ}$C) dell'ultimo mese, ottenendo questi dati ordinati $x_{(1)}, x_{(2)}, \ldots, x_{(30)}$:

10.2	10.7	11.5	12.3	12.6	13.1	13.7	13.8	13.8	13.9
14.2	14.2	14.5	14.7	14.7	15.1	16.8	17.2	17.3	17.5
17.6	17.8	18.0	18.1	18.3	19.4	19.6	25.3	26.4	31.0

Classi	FA		
$(10, 14]$	10		
$(14, 18]$	13		
$(18, 22]$	4		
$(22, 26]$			
$(26, 30]$			
$(30, 34]$			

Dati continui

ESEMPIO: Abbiamo misurato le temperature massime giornaliere (in °C) dell'ultimo mese, ottenendo questi dati ordinati $x_{(1)}, x_{(2)}, \ldots, x_{(30)}$:

10.2	10.7	11.5	12.3	12.6	13.1	13.7	13.8	13.8	13.9
14.2	14.2	14.5	14.7	14.7	15.1	16.8	17.2	17.3	17.5
17.6	17.8	18.0	18.1	18.3	19.4	19.6	25.3	26.4	31.0

Classi	FA		
$(10, 14]$	10		
$(14, 18]$	13		
$(18, 22]$	4		
$(22, 26]$	1		
$(26, 30]$			
$(30, 34]$			

Dati continui

ESEMPIO: Abbiamo misurato le temperature massime giornaliere (in °C) dell'ultimo mese, ottenendo questi dati ordinati $x_{(1)}, x_{(2)}, \ldots, x_{(30)}$:

10.2	10.7	11.5	12.3	12.6	13.1	13.7	13.8	13.8	13.9
14.2	14.2	14.5	14.7	14.7	15.1	16.8	17.2	17.3	17.5
17.6	17.8	18.0	18.1	18.3	19.4	19.6	25.3	26.4	31.0

Classi	FA		
$(10, 14]$	10		
$(14, 18]$	13		
$(18, 22]$	4		
$(22, 26]$	1		
$(26, 30]$	1		
$(30, 34]$			

Dati continui

ESEMPIO: Abbiamo misurato le temperature massime giornaliere (in °C) dell'ultimo mese, ottenendo questi dati ordinati $x_{(1)}, x_{(2)}, \ldots, x_{(30)}$:

10.2	10.7	11.5	12.3	12.6	13.1	13.7	13.8	13.8	13.9
14.2	14.2	14.5	14.7	14.7	15.1	16.8	17.2	17.3	17.5
17.6	17.8	18.0	18.1	18.3	19.4	19.6	25.3	26.4	31.0

Classi	FA		
$(10, 14]$	10		
$(14, 18]$	13		
$(18, 22]$	4		
$(22, 26]$	1		
$(26, 30]$	1		
$(30, 34]$	1		

Dati continui

ESEMPIO: Abbiamo misurato le temperature massime giornaliere (in °C) dell'ultimo mese, ottenendo questi dati ordinati $x_{(1)}, x_{(2)}, \ldots, x_{(30)}$:

10.2	10.7	11.5	12.3	12.6	13.1	13.7	13.8	13.8	13.9
14.2	14.2	14.5	14.7	14.7	15.1	16.8	17.2	17.3	17.5
17.6	17.8	18.0	18.1	18.3	19.4	19.6	25.3	26.4	31.0

Classi	FA	FR	
$(10, 14]$	10	$10/30 \simeq 0.33$	
$(14, 18]$	13	$13/30 \simeq 0.43$	
$(18, 22]$	4	$4/30 \simeq 0.13$	
$(22, 26]$	1	$1/30 \simeq 0.03$	
$(26, 30]$	1	$1/30 \simeq 0.03$	
$(30, 34]$	1	$1/30 \simeq 0.03$	

ESEMPIO: Abbiamo misurato le temperature massime giornaliere (in $^\circ$C) dell'ultimo mese, ottenendo questi dati ordinati $x_{(1)}, x_{(2)}, \ldots, x_{(30)}$:

10.2	10.7	11.5	12.3	12.6	13.1	13.7	13.8	13.8	13.9
14.2	14.2	14.5	14.7	14.7	15.1	16.8	17.2	17.3	17.5
17.6	17.8	18.0	18.1	18.3	19.4	19.6	25.3	26.4	31.0

Classi	FA	FR	FC
$(10, 14]$	10	$10/30 \simeq 0.33$	0.33
$(14, 18]$	13	$13/30 \simeq 0.43$	$0.33 + 0.43 = 0.76$
$(18, 22]$	4	$4/30 \simeq 0.13$	$0.76 + 0.13 = 0.89$
$(22, 26]$	1	$1/30 \simeq 0.03$	$0.89 + 0.03 = 0.92$
$(26, 30]$	1	$1/30 \simeq 0.03$	$0.92 + 0.03 = 0.95$
$(30, 34]$	1	$1/30 \simeq 0.03$	$0.95 + 0.03 = 0.98$

Dati continui

ESEMPIO: Abbiamo misurato le temperature massime giornaliere (in °C) dell'ultimo mese, ottenendo questi dati ordinati $x_{(1)}, x_{(2)}, \ldots, x_{(30)}$:

10.2	10.7	11.5	12.3	12.6	13.1	13.7	13.8	13.8	13.9
14.2	14.2	14.5	14.7	14.7	15.1	16.8	17.2	17.3	17.5
17.6	17.8	18.0	18.1	18.3	19.4	19.6	25.3	26.4	31.0

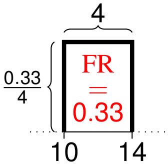
Classi	FA	FR	FC
$(10, 14]$	10	$10/30 \simeq 0.33$	0.33
$(14, 18]$	13	$13/30 \simeq 0.43$	$0.33 + 0.43 = 0.76$
$(18, 22]$	4	$4/30 \simeq 0.14$	$0.76 + 0.14 = 0.90$
$(22, 26]$	1	$1/30 \simeq 0.04$	$0.90 + 0.04 = 0.94$
$(26, 30]$	1	$1/30 \simeq 0.03$	$0.94 + 0.03 = 0.97$
$(30, 34]$	1	$1/30 \simeq 0.03$	$0.97 + 0.03 = 1.00$

Dati continui

ESEMPIO: Abbiamo misurato le temperature massime giornaliere (in $^\circ$C) dell'ultimo mese, ottenendo questi dati ordinati $x_{(1)}, x_{(2)}, \ldots, x_{(30)}$:

10.2	10.7	11.5	12.3	12.6	13.1	13.7	13.8	13.8	13.9
14.2	14.2	14.5	14.7	14.7	15.1	16.8	17.2	17.3	17.5
17.6	17.8	18.0	18.1	18.3	19.4	19.6	25.3	26.4	31.0

Classi	FA	FR	FC
$(10, 14]$	10	$10/30 \simeq 0.33$	0.33
$(14, 18]$	13	$13/30 \simeq 0.43$	$0.33 + 0.43 = 0.76$
$(18, 22]$	4	$4/30 \simeq 0.14$	$0.76 + 0.14 = 0.90$
$(22, 26]$	1	$1/30 \simeq 0.04$	$0.90 + 0.04 = 0.94$
$(26, 30]$	1	$1/30 \simeq 0.03$	$0.94 + 0.03 = 0.97$
$(30, 34]$	1	$1/30 \simeq 0.03$	$0.97 + 0.03 = 1.00$



$$\text{DENSITÀ}\,(C) = \frac{\mathrm{FR}(C)}{\text{ampiezza}\,(C)}$$

Dati continui

ESEMPIO: Abbiamo misurato le temperature massime giornaliere (in $^\circ$C) dell'ultimo mese, ottenendo questi dati ordinati $x_{(1)}, x_{(2)}, \ldots, x_{(30)}$:

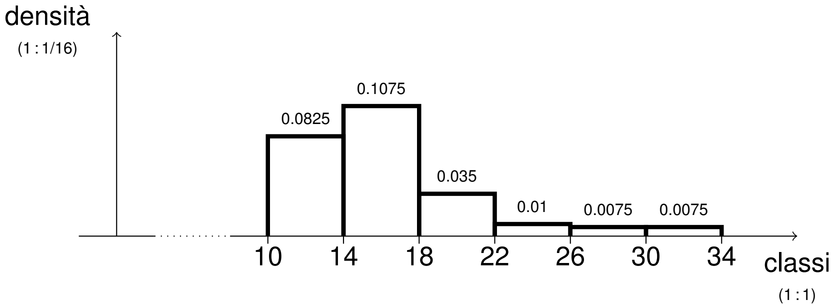
10.2	10.7	11.5	12.3	12.6	13.1	13.7	13.8	13.8	13.9
14.2	14.2	14.5	14.7	14.7	15.1	16.8	17.2	17.3	17.5
17.6	17.8	18.0	18.1	18.3	19.4	19.6	25.3	26.4	31.0

Classi	FA	FR	FC	Densità
$(10, 14]$	10	$10/30 \simeq 0.33$	0.33	$0.33/4 \simeq 0.0825$
$(14, 18]$	13	$13/30 \simeq 0.43$	$0.33 + 0.43 = 0.76$	$0.43/4 \simeq 0.1075$
$(18, 22]$	4	$4/30 \simeq 0.14$	$0.76 + 0.14 = 0.90$	$0.14/4 \simeq 0.035$
$(22, 26]$	1	$1/30 \simeq 0.04$	$0.90 + 0.04 = 0.94$	$0.04/4 \simeq 0.01$
$(26, 30]$	1	$1/30 \simeq 0.03$	$0.94 + 0.03 = 0.97$	$0.03/4 \simeq 0.0075$
$(30, 34]$	1	$1/30 \simeq 0.03$	$0.97 + 0.03 = 1.00$	$0.03/4 \simeq 0.0075$

$$\text{DENSITÀ}\,(C) = \frac{\mathrm{FR}(C)}{\text{ampiezza}\,(C)}$$

Dati continui

Classi	FA	FR	FC	Densità
$(10, 14]$	10	0.33	0.33	0.0825
$(14, 18]$	13	0.43	0.76	0.1075
$(18, 22]$	4	0.14	0.90	0.035
$(22, 26]$	1	0.04	0.94	0.01
$(26, 30]$	1	0.03	0.97	0.0075
$(30, 34]$	1	0.03	1.00	0.0075



Dati continui

Classi	FA	FR	FC	Densità
$(10, 14]$	10	0.33	0.33	0.0825
$(14, 18]$	13	0.43	0.76	0.1075
$(18, 22]$	4	0.14	0.90	0.035
$(22, 26]$	1	0.04	0.94	0.01
$(26, 30]$	1	0.03	0.97	0.0075
$(30, 34]$	1	0.03	1.00	0.0075

$$\bar{x} \simeq \sum_{\text{classi } C} \text{punto medio}\,(C) \cdot \mathrm{FR}(C)$$

Classi	FA	FR	FC	Densità
$(10, 14]$	10	0.33	0.33	0.0825
$(14, 18]$	13	0.43	0.76	0.1075
$(18, 22]$	4	0.14	0.90	0.035
$(22, 26]$	1	0.04	0.94	0.01
$(26, 30]$	1	0.03	0.97	0.0075
$(30, 34]$	1	0.03	1.00	0.0075

$$\bar{x} \simeq \sum_{\text{classi } C} \text{punto medio}\,(C) \cdot \mathrm{FR}(C)$$

$$= \frac{10+14}{2} \cdot 0.33 + \frac{14+18}{2} \cdot 0.43 + \ldots + \frac{30+34}{2} \cdot 0.03$$

$$= 16.4$$

da confrontarsi con

$$\bar{x} = 16.44\overline{3}$$

Dati continui

Classi	FA	FR	FC	Densità
$(10, 14]$	10	0.33	0.33	0.0825
$(14, 18]$	13	0.43	0.76	0.1075
$(18, 22]$	4	0.14	0.90	0.035
$(22, 26]$	1	0.04	0.94	0.01
$(26, 30]$	1	0.03	0.97	0.0075
$(30, 34]$	1	0.03	1.00	0.0075

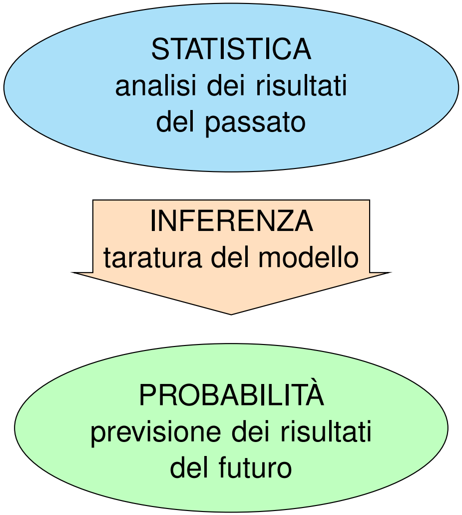
$$\left.\begin{array}{l} \mathrm{FC}(14) = 0.33 < 0.50 \;\Rightarrow\; m \text{ non può essere} \leq 14 \\ 1 - \mathrm{FC}(18) = 0.24 < 0.50 \;\Rightarrow\; m \text{ non può essere} > 18 \end{array}\right\} \Rightarrow$$

$$\Rightarrow \quad m \in (14, 18]$$

Programma

- Statistica descrittiva
  (riassumere molti dati attraverso poche caratteristiche essenziali)
- Probabilità
  (costruire un modello che preveda il risultato di un esperimento)
- Inferenza statistica
  (tarare i parametri del modello in base ai risultati dell'esperimento)
- Regressione lineare
  (riconoscere relazioni tra dati di tipo diverso)

Statistica e Probabilità



Esperimenti aleatori

ESPERIMENTO ALEATORIO $=$ esito non scontato

ESEMPI:

- lancio di un dado
- lancio di tre monete
- sondaggio tra 100 studenti

Esperimenti aleatori

ESPERIMENTO ALEATORIO $=$ esito non scontato

ESEMPI:

- lancio di un dado
- lancio di tre monete
- sondaggio tra 100 studenti

EVENTO $=$ proposizione circa il risultato dell'esperimento

ESEMPI:

- $E =$ "uscirà 6"
- $F =$ "uscirà testa al 2$^{\mathrm{o}}$ lancio"
- $G =$ "tutti e 100 gli intervistati saranno più bassi di $2\,\mathrm{m}$"

Esperimenti aleatori

$\mathcal{E} =$ insieme di tutti i possibili eventi

L'insieme $\mathcal{E}$ è una logica booleana con le operazioni

$$\wedge = \text{AND} \qquad \vee = \text{OR} \qquad \overline{\cdot} = \text{NOT}$$

Esperimenti aleatori

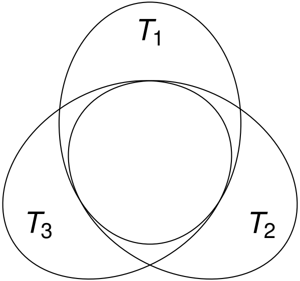
$\mathcal{E} =$ insieme di tutti i possibili eventi

L'insieme $\mathcal{E}$ è una logica booleana con le operazioni

$$\wedge = \text{AND} \qquad \vee = \text{OR} \qquad \overline{\cdot} = \text{NOT}$$

ESEMPIO: nel lancio di tre monete

$T_i =$ "uscirà testa all'i-esimo lancio" $\quad (i = 1, 2, 3)$



Esperimenti aleatori

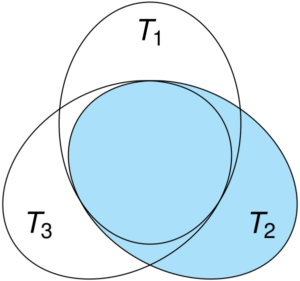
$\mathcal{E} =$ insieme di tutti i possibili eventi

L'insieme $\mathcal{E}$ è una logica booleana con le operazioni

$$\wedge = \text{AND} \qquad \vee = \text{OR} \qquad \overline{\cdot} = \text{NOT}$$

ESEMPIO: nel lancio di tre monete

$T_i =$ "uscirà testa all'i-esimo lancio" $\quad (i = 1, 2, 3)$



$=$ "uscirà testa al 2º lancio"

Esperimenti aleatori

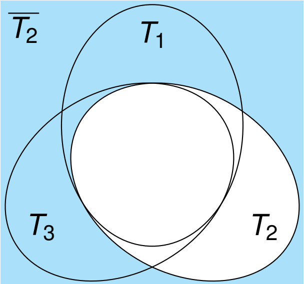
$\mathcal{E} =$ insieme di tutti i possibili eventi

L'insieme $\mathcal{E}$ è una logica booleana con le operazioni

$$\wedge = \text{AND} \qquad \vee = \text{OR} \qquad \overline{\cdot} = \text{NOT}$$

ESEMPIO: nel lancio di tre monete

$T_i =$ "uscirà testa all'i-esimo lancio" $\quad (i = 1, 2, 3)$



$=$ "uscirà croce al 2º lancio"

Esperimenti aleatori

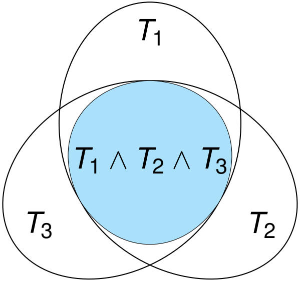
$\mathcal{E} =$ insieme di tutti i possibili eventi

L'insieme $\mathcal{E}$ è una logica booleana con le operazioni

$$\wedge = \text{AND} \qquad \vee = \text{OR} \qquad \overline{\cdot} = \text{NOT}$$

ESEMPIO: nel lancio di tre monete

$T_i =$ "uscirà testa all'i-esimo lancio" $\quad (i = 1, 2, 3)$



$=$ "uscirà sempre testa"

Esperimenti aleatori

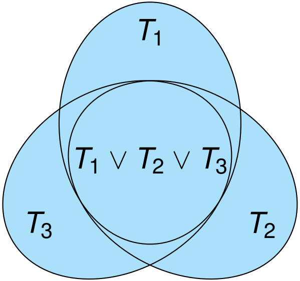
$\mathcal{E} =$ insieme di tutti i possibili eventi

L'insieme $\mathcal{E}$ è una logica booleana con le operazioni

$$\wedge = \text{AND} \qquad \vee = \text{OR} \qquad \overline{\cdot} = \text{NOT}$$

ESEMPIO: nel lancio di tre monete

$T_i =$ "uscirà testa all'i-esimo lancio" $\quad (i = 1, 2, 3)$



$=$ "almeno una volta uscirà testa"

Esperimenti aleatori

$\mathcal{E} =$ insieme di tutti i possibili eventi

L'insieme $\mathcal{E}$ è una logica booleana con le operazioni

$$\wedge = \text{AND} \qquad \vee = \text{OR} \qquad \overline{\cdot} = \text{NOT}$$

PROPRIETÀ:

$(E \wedge F) \wedge G = E \wedge (F \wedge G)$ e similmente con $\vee \leftrightarrow \wedge$

$\overline{E \wedge F} = \overline{E} \vee \overline{F}$ e similmente con $\vee \leftrightarrow \wedge$

$E \wedge (F \vee G) = (E \wedge F) \vee (E \wedge G)$ e similmente con $\vee \leftrightarrow \wedge$

Esperimenti aleatori

$\mathcal{E} =$ insieme di tutti i possibili eventi

L'insieme $\mathcal{E}$ è una logica booleana con le operazioni

$$\wedge = \text{AND} \qquad \vee = \text{OR} \qquad \overline{\cdot} = \text{NOT}$$

PROPRIETÀ:

$(E \wedge F) \wedge G = E \wedge (F \wedge G)$ e similmente con $\vee \leftrightarrow \wedge$

$\overline{E \wedge F} = \overline{E} \vee \overline{F}$ e similmente con $\vee \leftrightarrow \wedge$

$E \wedge (F \vee G) = (E \wedge F) \vee (E \wedge G)$ e similmente con $\vee \leftrightarrow \wedge$

Si definiscono inoltre

$$\Omega = \text{evento certo} \quad := E \vee \overline{E} \quad \forall E$$

$$\varnothing = \text{evento impossibile} \quad := E \wedge \overline{E} \quad \forall E$$

E ed F sono *incompatibili* quando $E \wedge F = \varnothing$

E *implica* F quando $E \wedge F = E$

Probabilità

Definizione

Una *probabilità* è una qualsiasi funzione $\mathbb{P} : \mathcal{E} \to \mathbb{R}$ t.c.

1. $\mathbb{P}(E) \geq 0 \quad \forall E$ (positività)
2. $\mathbb{P}(\Omega) = 1$ (normalizzazione)
3. Se $E_1, E_2, \ldots, E_n$ sono a due a due incompatibili,

   $\mathbb{P}(E_1 \vee E_2 \vee \ldots \vee E_n) = \mathbb{P}(E_1) + \mathbb{P}(E_2) + \ldots + \mathbb{P}(E_n)$ (additività)

Probabilità

Definizione

Una *probabilità* è una qualsiasi funzione $\mathbb{P} : \mathcal{E} \to \mathbb{R}$ t.c.

1. $\mathbb{P}(E) \geq 0 \quad \forall E$ (positività)
2. $\mathbb{P}(\Omega) = 1$ (normalizzazione)
3. Se $E_1, E_2, \ldots, E_n$ sono a due a due incompatibili,

   $\mathbb{P}(E_1 \vee E_2 \vee \ldots \vee E_n) = \mathbb{P}(E_1) + \mathbb{P}(E_2) + \ldots + \mathbb{P}(E_n)$ (additività)

Conseguenze

(a) $\mathbb{P}(\overline{E}) = 1 - \mathbb{P}(E)$

(b) $0 \leq \mathbb{P}(E) \leq 1$

(c) Se E implica F, allora $\mathbb{P}(E) \leq \mathbb{P}(F)$

Probabilità

Definizione

Una *probabilità* è una qualsiasi funzione $\mathbb{P} : \mathcal{E} \to \mathbb{R}$ t.c.

1. $\mathbb{P}(E) \geq 0 \quad \forall E$ (positività)
2. $\mathbb{P}(\Omega) = 1$ (normalizzazione)
3. Se $E_1, E_2, \ldots, E_n$ sono a due a due incompatibili,

   $\mathbb{P}(E_1 \vee E_2 \vee \ldots \vee E_n) = \mathbb{P}(E_1) + \mathbb{P}(E_2) + \ldots + \mathbb{P}(E_n)$ (additività)

ATTENZIONE:

$$\mathbb{P}(T_1 \vee T_2 \vee T_3) \neq \mathbb{P}(T_1) + \mathbb{P}(T_2) + \mathbb{P}(T_3) = 150\% > 1$$

Probabilità

(a) $\mathbb{P}\left(\overline{E}\right) = 1 - \mathbb{P}(E)$:

$$1 \overset{(2)}{=} \mathbb{P}(\Omega)$$

Probabilità

(a) $\mathbb{P}\left(\overline{E}\right) = 1 - \mathbb{P}(E)$:

$$1 = \mathbb{P}(\Omega) = \mathbb{P}\left(E \vee \overline{E}\right)$$

Probabilità

(a) $\mathbb{P}\left(\overline{E}\right) = 1 - \mathbb{P}(E)$:

$$1 = \mathbb{P}(\Omega) = \mathbb{P}\left(E \vee \overline{E}\right) \overset{(3)}{=} \mathbb{P}(E) + \mathbb{P}\left(\overline{E}\right) \qquad \text{perché } E \wedge \overline{E} = \varnothing$$

Probabilità

(a) $\mathbb{P}\left(\overline{E}\right) = 1 - \mathbb{P}(E)$:

$$1 \;=\; \mathbb{P}(\Omega) \;= \mathbb{P}\left(E \vee \overline{E}\right) \;=\; \mathbb{P}(E) + \mathbb{P}\left(\overline{E}\right) \qquad \text{perché } E \wedge \overline{E} = \varnothing$$
$$\Rightarrow \quad \mathbb{P}\left(\overline{E}\right) = 1 - \mathbb{P}(E)$$

Probabilità

(a) $\mathbb{P}\left(\overline{E}\right) = 1 - \mathbb{P}\left(E\right)$:

$$1 \; = \; \mathbb{P}\left(\Omega\right) \; = \mathbb{P}\left(E \vee \overline{E}\right) \; = \; \mathbb{P}\left(E\right) + \mathbb{P}\left(\overline{E}\right) \qquad \text{perché } E \wedge \overline{E} = \varnothing$$

$$\Rightarrow \quad \mathbb{P}\left(\overline{E}\right) = 1 - \mathbb{P}\left(E\right)$$

(b) $0 \leq \mathbb{P}\left(E\right) \leq 1$:

$$0 \overset{(1)}{\leq} \mathbb{P}\left(E\right)$$

Probabilità

(a) $\mathbb{P}\left(\overline{E}\right) = 1 - \mathbb{P}(E)$:

$$1 \;=\; \mathbb{P}(\Omega) \;=\mathbb{P}\left(E \vee \overline{E}\right) \;=\; \mathbb{P}(E) + \mathbb{P}\left(\overline{E}\right) \qquad \text{perché } E \wedge \overline{E} = \varnothing$$
$$\Rightarrow \quad \mathbb{P}\left(\overline{E}\right) = 1 - \mathbb{P}(E)$$

(b) $0 \leq \mathbb{P}(E) \leq 1$:

$$0 \;\leq\; \mathbb{P}(E) \;= 1 - \mathbb{P}\left(\overline{E}\right)$$

Probabilità

(a) $\mathbb{P}\left(\overline{E}\right) = 1 - \mathbb{P}(E)$:

$$1 = \mathbb{P}(\Omega) = \mathbb{P}\left(E \vee \overline{E}\right) = \mathbb{P}(E) + \mathbb{P}\left(\overline{E}\right) \qquad \text{perché } E \wedge \overline{E} = \varnothing$$

$$\Rightarrow \quad \mathbb{P}\left(\overline{E}\right) = 1 - \mathbb{P}(E)$$

(b) $0 \leq \mathbb{P}(E) \leq 1$:

$$0 \leq \mathbb{P}(E) = 1 - \underset{\substack{|\vee\,(1) \\ 0}}{\mathbb{P}\left(\overline{E}\right)} \leq 1$$

Probabilità

(a) $\mathbb{P}\left(\overline{E}\right) = 1 - \mathbb{P}(E)$:

$$1 \;=\; \mathbb{P}(\Omega) \;= \mathbb{P}\left(E \vee \overline{E}\right) \;=\; \mathbb{P}(E) + \mathbb{P}\left(\overline{E}\right) \qquad \text{perché } E \wedge \overline{E} = \varnothing$$
$$\Rightarrow \quad \mathbb{P}\left(\overline{E}\right) = 1 - \mathbb{P}(E)$$

(b) $0 \leq \mathbb{P}(E) \leq 1$:

$$0 \;\leq\; \mathbb{P}(E) \;= 1 - \mathbb{P}\left(\overline{E}\right) \;\leq 1$$

(c) Se E implica F, allora $\mathbb{P}(E) \leq \mathbb{P}(F)$:

$$\mathbb{P}(F) = \mathbb{P}\left((E \wedge F) \vee (\overline{E} \wedge F)\right)$$

(a) $\mathbb{P}\left(\overline{E}\right) = 1 - \mathbb{P}(E)$:

$$1 = \mathbb{P}(\Omega) = \mathbb{P}\left(E \vee \overline{E}\right) = \mathbb{P}(E) + \mathbb{P}\left(\overline{E}\right) \qquad \text{perché } E \wedge \overline{E} = \varnothing$$
$$\Rightarrow \quad \mathbb{P}\left(\overline{E}\right) = 1 - \mathbb{P}(E)$$

(b) $0 \leq \mathbb{P}(E) \leq 1$:

$$0 \leq \mathbb{P}(E) = 1 - \mathbb{P}\left(\overline{E}\right) \leq 1$$

(c) Se E implica F, allora $\mathbb{P}(E) \leq \mathbb{P}(F)$:

$$\mathbb{P}(F) = \mathbb{P}\left((E \wedge F) \vee (\overline{E} \wedge F)\right)$$
$$\stackrel{(3)}{=} \mathbb{P}(E \wedge F) + \mathbb{P}\left(\overline{E} \wedge F\right) \qquad \text{perché } (E \wedge F) \wedge (\overline{E} \wedge F) = \varnothing$$

(a) $\mathbb{P}\left(\overline{E}\right) = 1 - \mathbb{P}(E):$

$$1 \;=\; \mathbb{P}(\Omega) = \mathbb{P}\left(E \vee \overline{E}\right) \;=\; \mathbb{P}(E) + \mathbb{P}\left(\overline{E}\right) \qquad \text{perché } E \wedge \overline{E} = \varnothing$$
$$\Rightarrow \quad \mathbb{P}\left(\overline{E}\right) = 1 - \mathbb{P}(E)$$

(b) $0 \leq \mathbb{P}(E) \leq 1:$

$$0 \leq \; \mathbb{P}(E) \; = 1 - \mathbb{P}\left(\overline{E}\right) \; \leq 1$$

(c) Se E implica F, allora $\mathbb{P}(E) \leq \mathbb{P}(F):$

$$\mathbb{P}(F) = \mathbb{P}\left((E \wedge F) \vee (\overline{E} \wedge F)\right)$$
$$= \; \mathbb{P}(E \wedge F) + \mathbb{P}\left(\overline{E} \wedge F\right)$$
$$= \mathbb{P}(E) + \mathbb{P}\left(\overline{E} \wedge F\right) \qquad \text{perché } E \wedge F = E$$

Probabilità

(a) $\mathbb{P}\left(\overline{E}\right) = 1 - \mathbb{P}(E)$:

$$1 = \mathbb{P}(\Omega) = \mathbb{P}\left(E \vee \overline{E}\right) = \mathbb{P}(E) + \mathbb{P}\left(\overline{E}\right) \qquad \text{perché } E \wedge \overline{E} = \varnothing$$

$$\Rightarrow \quad \mathbb{P}\left(\overline{E}\right) = 1 - \mathbb{P}(E)$$

(b) $0 \leq \mathbb{P}(E) \leq 1$:

$$0 \leq \mathbb{P}(E) = 1 - \mathbb{P}\left(\overline{E}\right) \leq 1$$

(c) Se E implica F, allora $\mathbb{P}(E) \leq \mathbb{P}(F)$:

$$\begin{aligned}\mathbb{P}(F) &= \mathbb{P}\left((E \wedge F) \vee (\overline{E} \wedge F)\right) \\ &= \mathbb{P}(E \wedge F) + \mathbb{P}\left(\overline{E} \wedge F\right) \\ &= \mathbb{P}(E) + \underset{\substack{|\vee (1) \\ 0}}{\mathbb{P}\left(\overline{E} \wedge F\right)} \\ &\geq \mathbb{P}(E)\end{aligned}$$