Indici di Dispersione

Si cercano indici di dispersione che:

- utilizzino tutti i dati $\{x_1, x_2, \ldots, x_n\}$
- siano basati sulla nozione di scarto (distanza) dei dati
  - rispetto a un centro $d_i = |x_i - C|$
    ad esempio, rispetto alla media aritmetica $d_i = |x_i - \bar{x}|$
  - rispetto a un dato $d_i = |x_i - x_j|$

con alcune proprietà generali:

- l'indice di dispersione non deve mai essere negativo
- assume il valore 0 se i dati sono tutti uguali
- non cambia se si aggiunge una costante ai dati

Varianza

Varianza: è la media aritmetica *(semplice o ponderata)* dei quadrati degli scarti.

- Dato l'insieme di valori $\{x_1, x_2, \ldots, x_n\}$

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

- Dato l'insieme di valori $\{x_1, x_2, \ldots, x_m\}$ con le rispettive frequenze assolute $\{f_1, f_2, \ldots, f_m\}$

$$\text{Var} = s^2 = \frac{1}{n} \cdot \sum_{i=1}^{m} f_i \cdot (x_i - \bar{x})^2 \qquad \text{dove } n = \sum_{i=1}^{m} f_i$$

Deviazione Standard

Deviazione standard *(o scarto quadratico medio)*: è la radice quadrata della varianza.

$$s = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2} \quad \text{oppure} \quad s = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{m} f_i \cdot (x_i - \bar{x})^2}$$

Consente di avere un indice di dispersione espresso nella stessa unità di misura dei dati.

Nota: applicando una trasformazione lineare ai dati

$$y_i = a x_i + b \qquad \Rightarrow \qquad s_y^2 = a^2 s_x^2 \,, \qquad s_y = |a|\, s_x$$

Statistiche Campionarie

Spesso gli *indici statistici* vengono applicati non all'intera *popolazione*, ma a un suo campione. Si cerca di stimare (inferenza) nel miglior modo possibile le caratteristiche dell'intera popolazione a partire dalle informazioni desunte da un campione rappresentativo.
In questo caso si utilizzano le seguenti formule modificate:

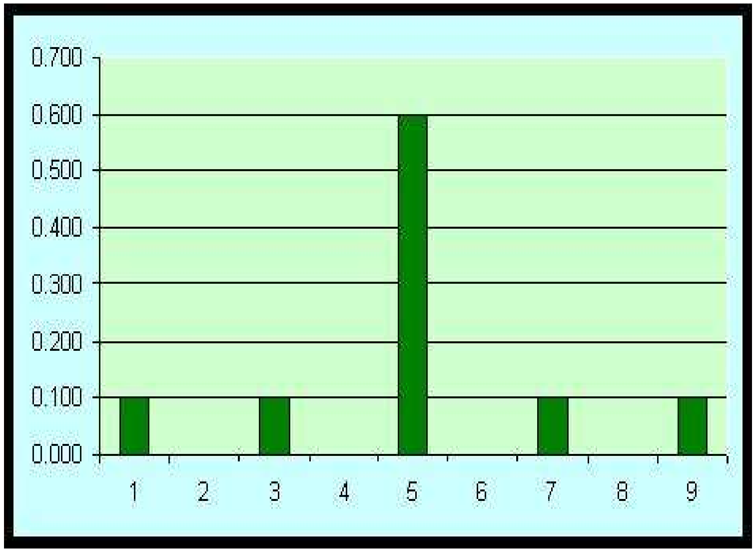
Varianza campionaria *(stimata)*:

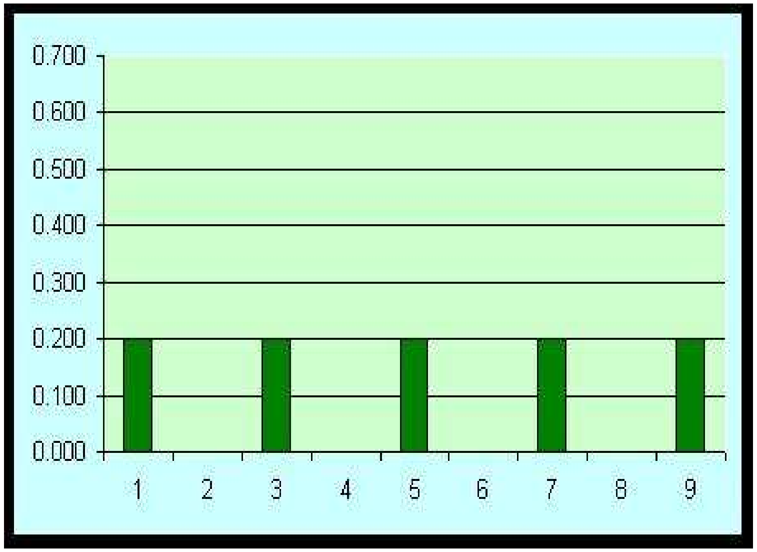
$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Deviazione standard campionaria *(stimata)*:

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

Esempio Riassuntivo





Esempio Riassuntivo

Caso A

x_i	f_i	f_i/n
1	1	0.100
3	1	0.100
5	6	0.600
7	1	0.100
9	1	0.100
	10	1.000

Caso B

x_i	f_i	f_i/n
1	2	0.200
3	2	0.200
5	2	0.200
7	2	0.200
9	2	0.200
	10	1.000

	Caso A		Caso B
media	5.00	media	5,00
mediana	5.00	mediana	5.00
varianza	4.00	varianza	8.00
varianza stimata	4.44	varianza stimata	8.89
deviazione standard	2.00	deviazione standard	2.83
deviazione standard stimata	2.11	deviazione standard stimata	2.98

Esercizi

Esercizio 1. Trovare media, mediana, moda, varianza e deviazione standard dei seguenti dati non ordinati e non raggruppati. Tracciare l'istogramma delle frequenze.

7 4 10 9 15 12 7 8 11 4 14 10 5 14 1 10 8 12 6 5

Esercizi

Soluzione: si costruisce la tabella della distribuzione di frequenza

x	1	4	5	6	7	8	9	10	11	12	14	15	
f_{ass}	1	2	2	1	2	2	1	3	1	2	2	1	**20**

$$\bar{x} = \frac{1}{20}(1+8+10+6+14+16+9+30+11+24+28+15) = 8.6$$

$$s^2 = \frac{1}{20}(57.76+42.32+25.92+6.76+5.12+0.72+0.16+5.88+$$
$$+5.76+23.12+58.33+40.96) \simeq 13.64$$

$$s \simeq 3.69$$

moda $= 10.0$ mediana $= 8.5$

Esercizi

Esercizio 2. Nel rilevare l'altezza in cm di un gruppo di reclute si è ottenuta la seguente tabella delle frequenze. Calcolare media, mediana e quartili.

Soluzione:

cm	f_{ass}	f_{cum}
166	1	1
168	3	4
169	6	10
170	11	21
171	8	29
172	6	35
173	4	39
174	3	42
175	1	43
178	1	44

$n = 44$ dimensione del campione

$\bar{x} \simeq 170.9$ media

$$M_e = \frac{x_{22} + x_{23}}{2} = 171 \quad \text{mediana}$$

$$q_1 = \frac{x_{11} + x_{12}}{2} = 170 \quad \text{primo quartile}$$

$$q_3 = \frac{x_{33} + x_{34}}{2} = 172 \quad \text{terzo quartile}$$

$q_3 - q_1 = 2$ distanza interquartile

La distanza interquartile è un altro indice di dispersione, legato alla nozione di mediana. La mediana suddivide l'insieme dei dati ordinati $\{x_i\}$ in due parti ugualmente numerose. I quartili si ottengono suddividendo i dati ordinati in quattro parti ugualmente numerose.

Esercizi

Esercizio 3. Si consideri la seguente tabella relativa alle frequenze dei pesi in Kg di 100 individui adulti.

Peso p in Kg	f_{ass}
$50 \leq p < 55$	20
$55 \leq p < 60$	15
$60 \leq p < 65$	18
$65 \leq p < 70$	22
$70 \leq p < 75$	18
$75 \leq p < 80$	7

- le classi sono di uguale ampiezza
- supponiamo che i dati siano uniformemente distribuiti all'interno di ogni classe
- possiamo definire per ogni classe un rappresentante r_i *(class mark)*

Esercizi

Soluzione: calcoliamo la media e lo scarto quadratico medio utilizzando i valori dei rappresentanti.

Peso p in Kg	f_i	F_i	r_i
$50 \leq p < 55$	20	20	52.5
$55 \leq p < 60$	15	35	57.5
$60 \leq p < 65$	18	53	62.5
$65 \leq p < 70$	22	75	67.5
$70 \leq p < 75$	18	93	72.5
$75 \leq p < 80$	7	100	77.5

Calcoliamo il peso medio:

$$\begin{aligned}\bar{p} &= \frac{1}{100}(20 \cdot 52.5 + 15 \cdot 57.5 + 18 \cdot 62.5 + 22 \cdot 67.5 + \\ &\quad + 18 \cdot 72.5 + 7 \cdot 77.5) = 63.7\end{aligned}$$

Esercizi

Calcoliamo la varianza e lo scarto quadratico medio:

r_i	$r_i - \bar{p}$	$(r_i - \bar{p})^2$	f_i
52.5	−11.2	125.44	20
57.5	−6.2	38.44	15
62.5	−1.2	1.44	18
67.5	3.8	14.44	22
72.5	8.8	77.44	18
77.5	13.8	190.44	7

$$\begin{aligned} s^2 &= \frac{1}{100}(20 \cdot 125.44 + 15 \cdot 38.44 + 18 \cdot 1.44 + 22 \cdot 14.44 + \\ &\quad +18 \cdot 77.44 + 7 \cdot 190.44) \simeq 61.56\,\mathrm{Kg}^2 \end{aligned}$$

$s \simeq 7.85\,\mathrm{Kg}$

Media – Varianza – Deviazione Standard

$\bar{x}$	media	$\frac{1}{n} \cdot \sum_{i=1}^{n} x_i$	$\frac{1}{n} \cdot \sum_{i=1}^{m} f_i \, x_i$
s^2	varianza	$\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$	$\frac{1}{n} \cdot \sum_{i=1}^{m} f_i \cdot (x_i - \bar{x})^2$
s	dev. standard	$\sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2}$	$\sqrt{\frac{1}{n} \cdot \sum_{i=1}^{m} f_i \cdot (x_i - \bar{x})^2}$
s^2	campionaria	$\frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$	$\frac{1}{n-1} \cdot \sum_{i=1}^{m} f_i \cdot (x_i - \bar{x})^2$
s^2	campionaria	$\sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2}$	$\sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{m} f_i \cdot (x_i - \bar{x})^2}$

Varianza – Deviazione Standard

Le espressioni della varianza (e della deviazione standard) possono essere riscritte come segue:

$$s^2 = \frac{1}{n} \cdot \left(\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2 \right) \qquad \text{o} \qquad s^2 = \frac{1}{n} \cdot \left(\sum_{i=1}^{m} f_i\, x_i^2 - n\,\bar{x}^2 \right)$$

Infatti,

$$\begin{aligned} \sum_{i=1}^{n} (x_i - \bar{x})^2 &= \sum_{i=1}^{n} (x_i^2 - 2x_i\,\bar{x} + \bar{x}^2) = \sum_{i=1}^{n} x_i^2 - 2\bar{x} \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \bar{x}^2 = \\ &= \sum_{i=1}^{n} x_i^2 - 2\bar{x}(n\,\bar{x}) + n\,\bar{x}^2 = \sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2 \end{aligned}$$

Esercizio

Un'indagine su un campione di $n = 100$ studenti, che hanno sostenuto la prova scritta di matematica, ha prodotto il seguente risultato. Le votazioni in centesimi sono state raggruppate in quattro *classi*.

classe	f_i	f_i/n
20 – 40	10	0.10
40 – 60	20	0.20
60 – 80	50	0.50
80 – 100	20	0.20
	100	1.00

Le classi sono di uguale ampiezza e contigue.

Esercizio

Nell'ipotesi di *distribuzione uniforme*, è naturale associare a ciascuna classe, come *rappresentante*, il valore centrale r_i della classe stessa.

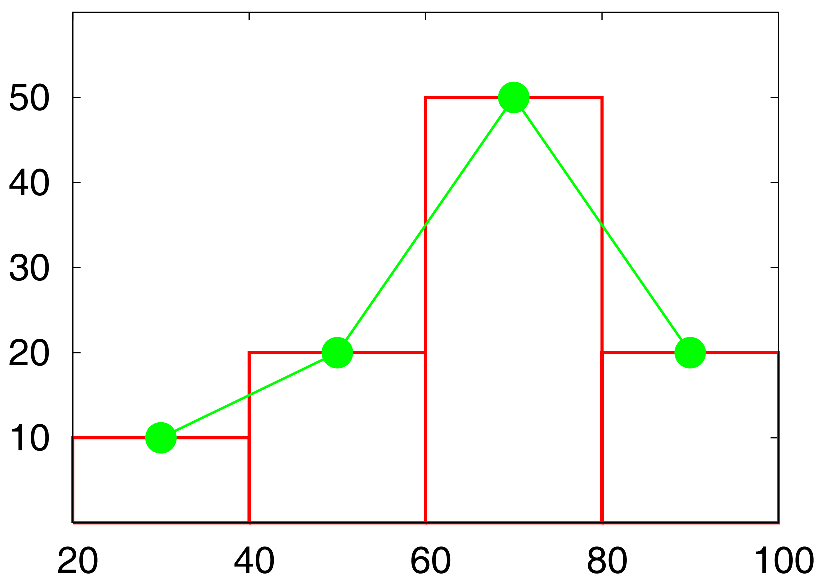
classe	r_i	f_i	F_i
$20-40$	30	10	10
$40-60$	50	20	30
$60-80$	70	50	80
$80-100$	90	20	100

$$\text{media} = \frac{1}{100} \cdot (10 \cdot 30 + 20 \cdot 50 + 50 \cdot 70 + 20 \cdot 90) = 66$$

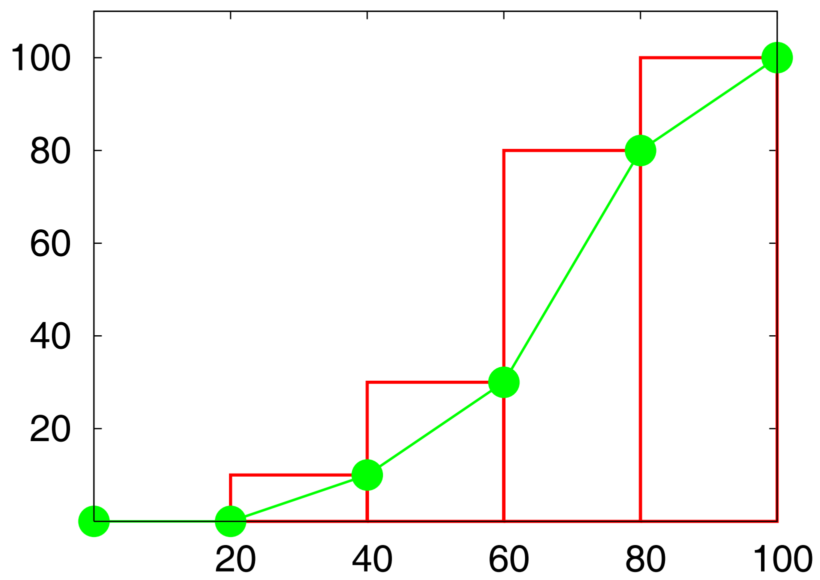
$\text{mediana} = 70$ (guardando solo i rappresentanti)

$$\text{varianza} = \frac{1}{100} \cdot (10 \cdot 36^2 + 20 \cdot 16^2 + 50 \cdot 4^2 + 20 \cdot 24^2) = 21.44$$

Esercizio



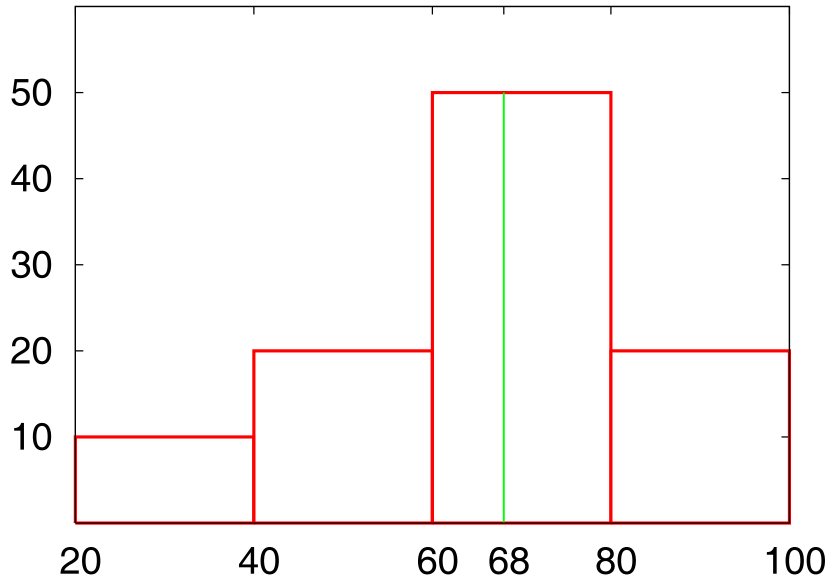
Poligono delle frequenze



Ogiva di frequenza

Esercizio

Calcolo della mediana:

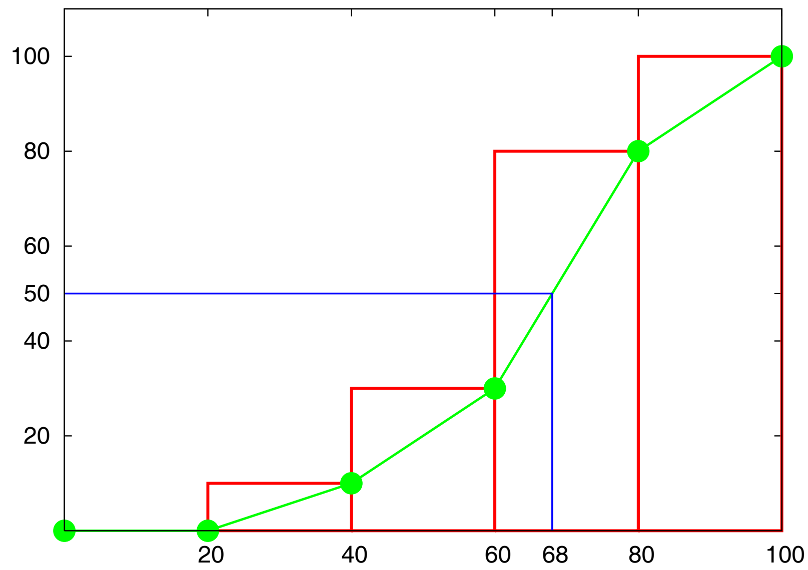


area totale $= 20 \cdot (10 + 20 + 50 + 20) = 2000$

Cerchiamo il valore $x = M_e$ tale che

$$20 \cdot 10 + 20 \cdot 20 + (x - 60) \cdot 50 = 1000 \quad \Rightarrow \quad x = 68 \quad \Rightarrow \quad M_e = 68$$

Esercizio



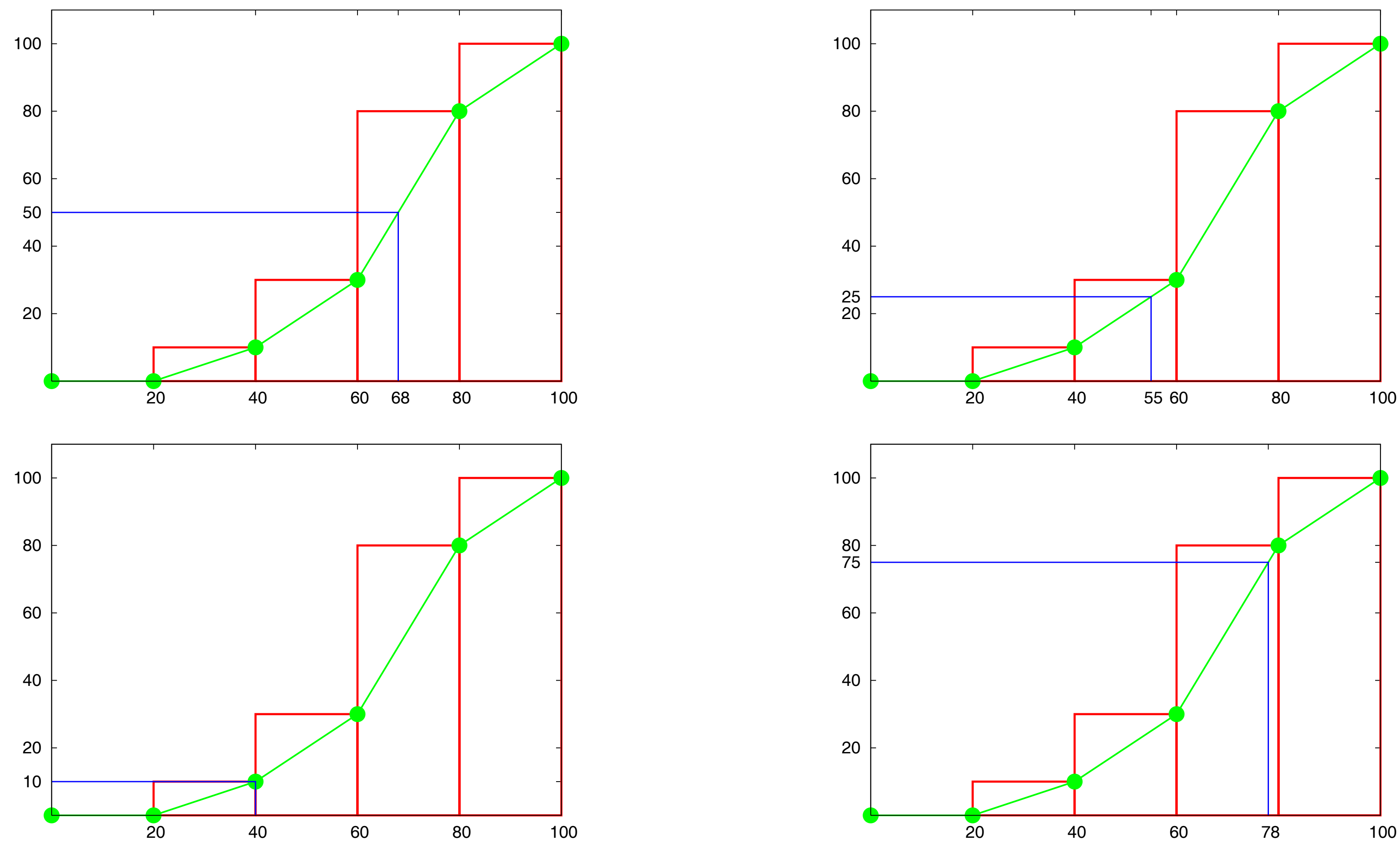
Calcolo della mediana: interpolazione lineare sui punti $A = (60, 30)$ e $B = (80, 80)$

$$\begin{cases} y = 50 \\ y - 30 = \frac{5}{2} \cdot (x - 60) \end{cases} \quad \Rightarrow \quad (50 - 30) = \frac{5}{2} \cdot (x - 60) \;\Rightarrow\; x = 68 \;\Rightarrow\; M_e = 68$$

Domande: (a) qual è il voto massimo preso dal primo 10% degli studenti (quelli che hanno preso il voto pìì basso)? [40]

(b) Calcolare i quartili dall'ogiva di frequenza. $[q_1 = 55\,,\ q_2 = 78]$

Esercizio


100
80
60
50
40
20
20 40 60 68 80 100
100
80
60
40
25
20
20 40 55 60 80 100
100
80
60
40
20
10
20 40 60 80 100
100
80
75
60
40
20
20 40 60 78 100