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,

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

Esercizi

Esercizio 1. 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.

Esercizio 2. 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 $= 8.5$

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.

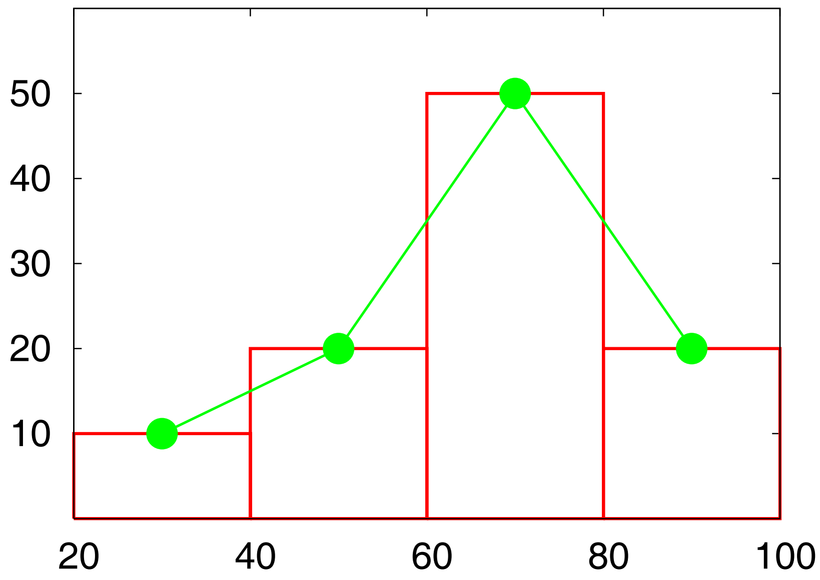
classe	r_i	f_i	F_i
20 – 40	30	10	10
40 – 60	50	20	30
60 – 80	70	50	80
80 – 100	90	20	100

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

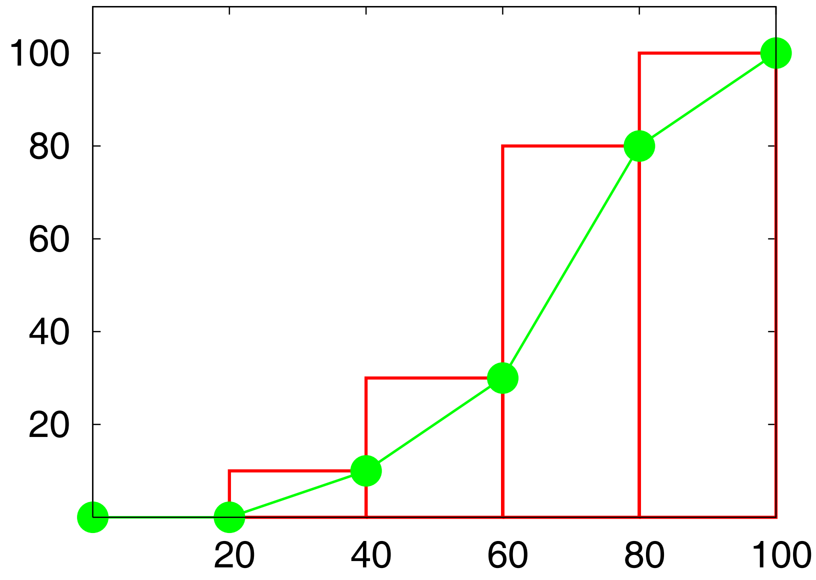
mediana $= 70$ (guardando solo i rappresentanti)

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

Esercizio



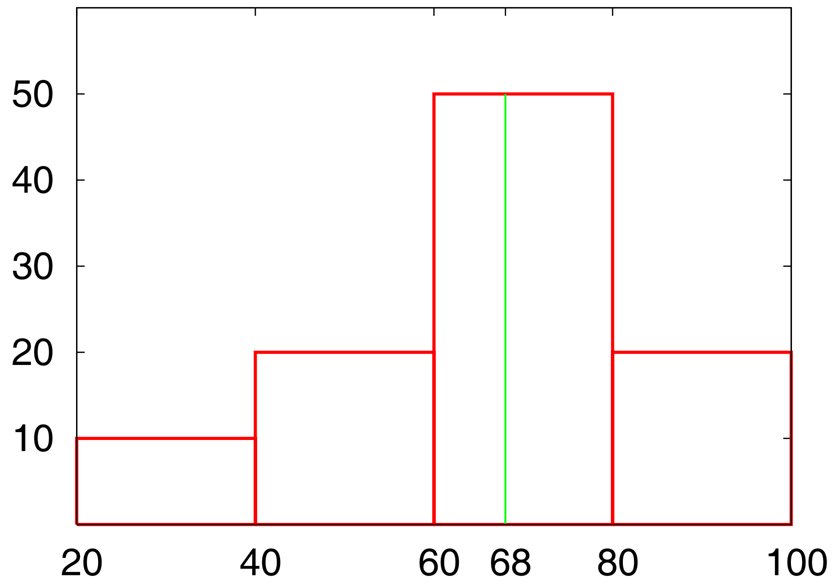
Poligono delle frequenze



Ogiva di frequenza

Esercizio

Calcolo della mediana:

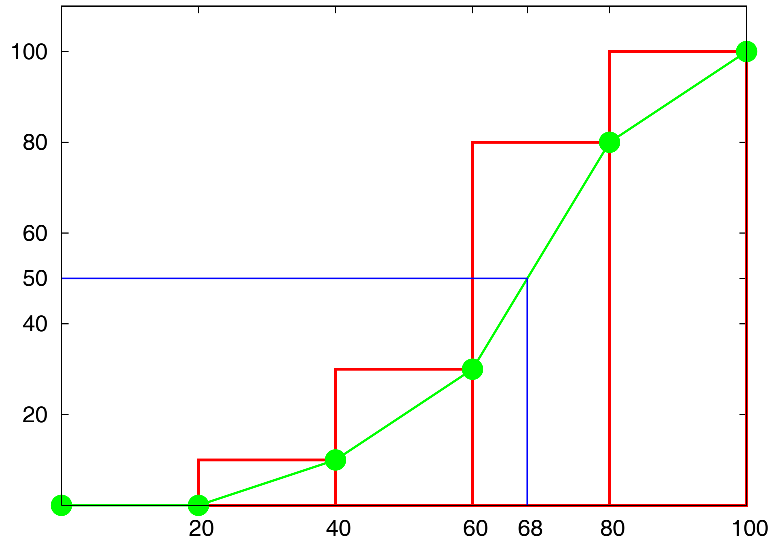


area totale $= 20 \cdot (10 + 20 + 50 + 20) = 2000$

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



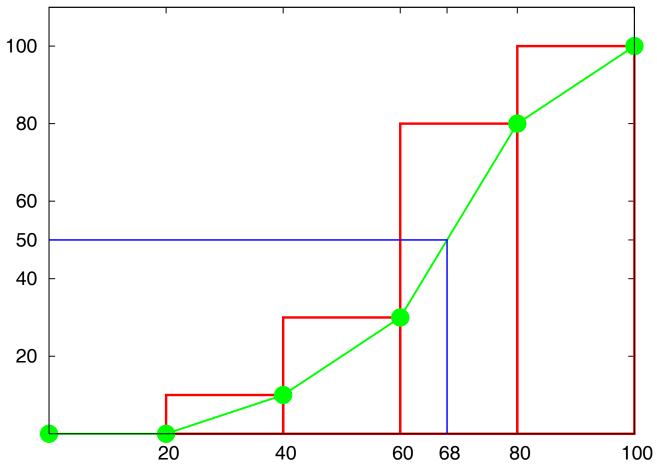
Calcolo della mediana: interpolazione lineare sui punti $A = (60, 30)$ e $B = (80, 80)$

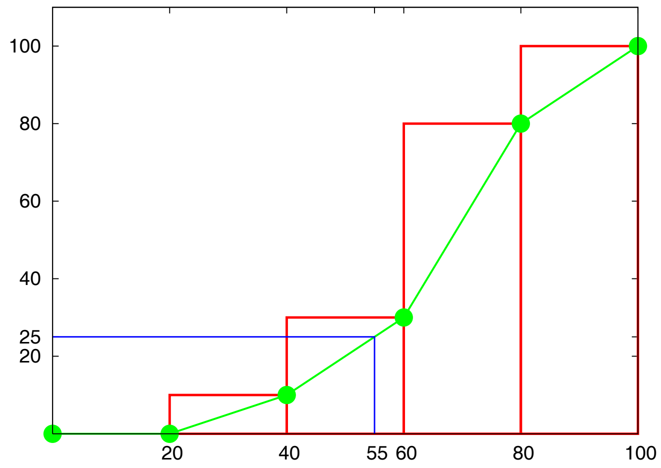
$$\begin{cases} y = 50 \\ y - 30 = \dfrac{5}{2} \cdot (x - 60) \end{cases} \Rightarrow (50 - 30) = \frac{5}{2} \cdot (x - 60) \Rightarrow x = 68 \Rightarrow M_e = 68$$

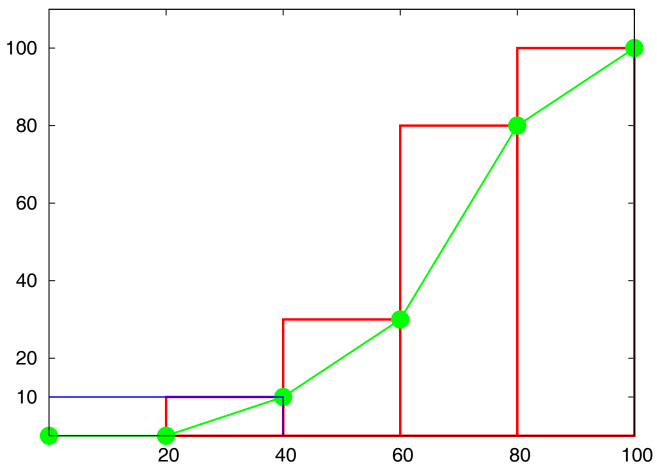
Domande: (a) qual è il voto massimo preso dal 10% degli studenti con i voti peggiori? [40]

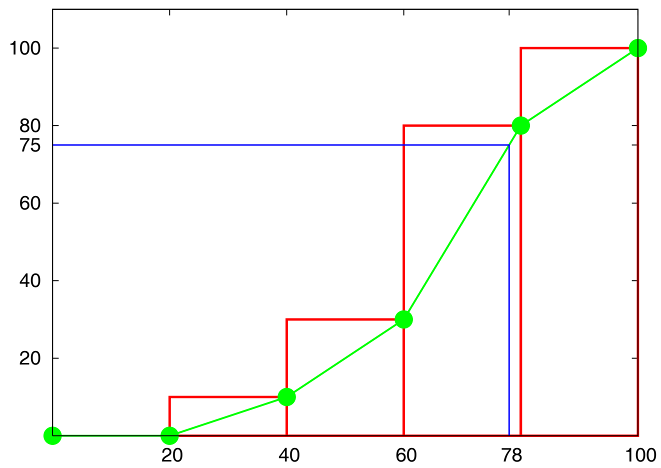
(b) Calcolare i quartili dall'ogiva di frequenza. $[q_1 = 55\,,\ q_2 = 78]$

Esercizio

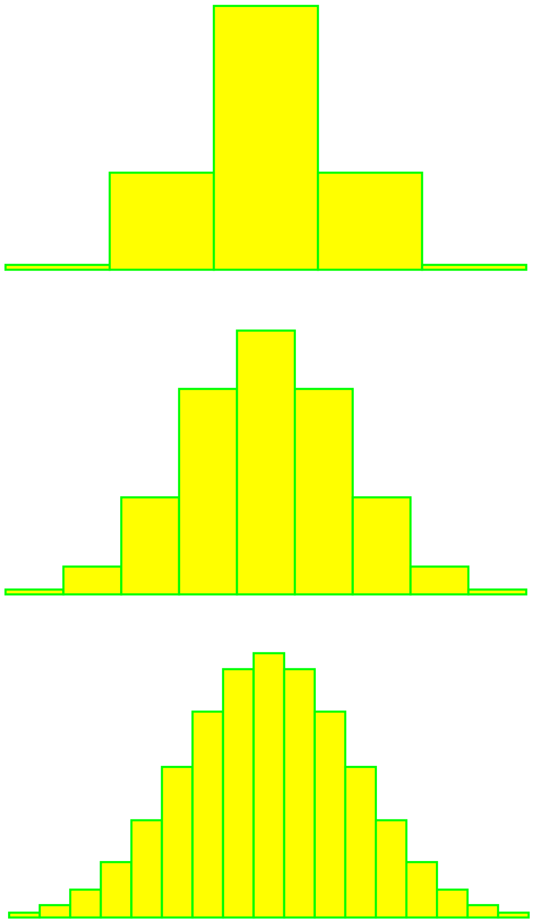

100
80
60
50
40
20
20
40
60
68
80
100


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

Distribuzione Normale

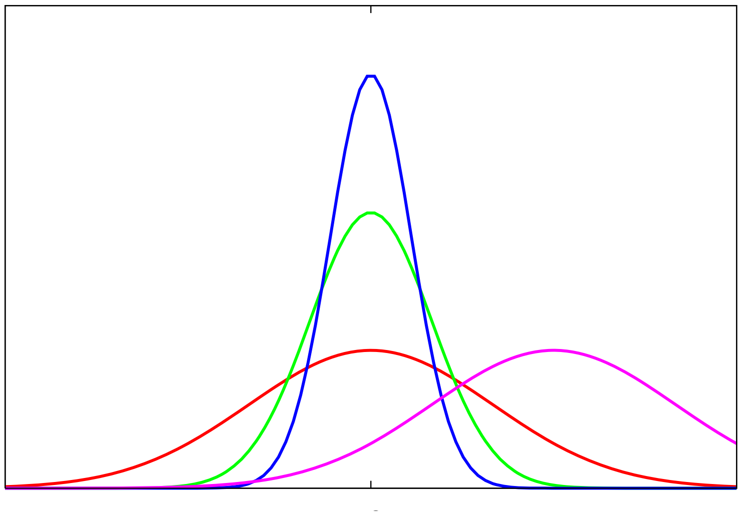


- istogramma delle frequenze di un insieme di misure relative a una grandezza che può variare con *continuità*
- popolazione molto numerosa, costituita da una quantità praticamente illimitata di individui *(popolazione infinita)*
- area dell'istogramma uguale a 1 *(normalizzata)*
- aumentando il numero di intervallini $n = 5, 9, 17, \ldots$ l'istogramma tende a stabilizzarsi intorno a una forma limite: *la curva di distribuzione delle frequenze*
- nel caso in figura: $y = Ae^{-B(x-C)^2}$ *distribuzione normale o gaussiana*

Distribuzione Normale

CURVE GAUSSIANE

$$y = Ae^{-B(x-C)^2}$$



Se la distribuzione è di tipo gaussiano con

- *media aritmetica* μ
- *deviazione standard* σ

si ha

$$A = \frac{1}{\sigma\sqrt{2\pi}} \quad B = \frac{1}{2\sigma^2} \quad C = \mu$$

La corrispondente curva normale sarà

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Curva normale standardizzata:

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad \mu = 0 \, , \sigma = 1$$