

Statistica

La statistica è la scienza che organizza e analizza dati numerici per fini descrittivi o per permettere di prendere delle decisioni e fare previsioni.

Statistica descrittiva: dalla mole di dati numerici a disposizione trae degli indicatori sintetici che possano riassumere le proprietà salienti dell'intera distribuzione.

Statistica inferenziale: utilizza dati statistici per previsioni di tipo probabilistico su situazioni future (*incerte*), su popolazioni più ampie . . .

Popolazione: *serie di dati*, che rappresenta l'insieme che si vuole indagare (reali, sperimentali, matematici)

Campione: *serie di dati*, che rappresenta una porzione della popolazione (campione rappresentativo)

Variabili: qualitative, quantitative (continue, discrete)

Distribuzione di Frequenza: Esempio

Supponiamo di avere un campione di $n = 200$ famiglie, di cui rileviamo il seguente carattere: **titolo di studio del capofamiglia**.

Questo carattere può presentare $m = 5$ differenti realizzazioni (*categorie*).

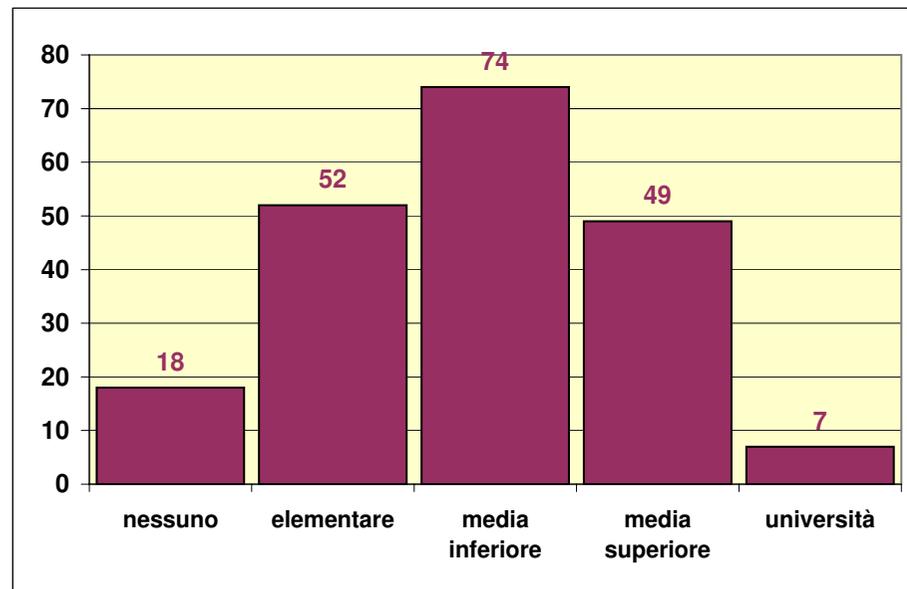
Costruiamo la tabella della **distribuzione di frequenza**:

	f_i	f_i/n	F_i	F_i/n
Nessun titolo	18	0.090	18	0.090
Licenza elementare	52	0.260	70	0.350
Diploma scuola media inferiore	74	0.370	144	0.720
Diploma scuola media superiore	49	0.245	193	0.965
Laurea	7	0.035	200	1.000
	200	1.000		

Distribuzione di Frequenza: Esempio

Rappresentiamo i dati riportati nella tabella della distribuzione delle frequenze con un **istogramma delle frequenze**.

titolo di studio	fi	fi / n
nessuno	18	0,09
elementare	52	0,260
media inferiore	74	0,370
media superiore	49	0,245
università	7	0,035
	200	1,000



- ogni rettangolo rappresenta un carattere
- l'area del rettangolo è proporzionale alla frequenza di quel carattere

Distribuzione di Frequenza

Dati raggruppati in **classi** o **categorie**: $(x_i, f_i)_{i=1, \dots, m}$

Frequenza assoluta f_i : è il numero di *osservazioni* che ricadono in ciascuna classe.

Il numero totale di osservazioni è $n = \sum_{i=1}^m f_i$.

Frequenza relativa f_i/n : è il rapporto tra la frequenza assoluta e il numero totale n di osservazioni. Rappresenta la percentuale di osservazioni in ogni classe o categoria.

Frequenza assoluta cumulata F_i :
$$F_i = \sum_{k=1}^i f_k$$

Frequenza relativa cumulata F_i/n :
$$\frac{1}{n} \sum_{k=1}^i f_k$$

Statistica Descrittiva

Misure, indici (numerici) che descrivono le caratteristiche della distribuzione di una o più variabili in modo sintetico.

- indici di posizione o centralità:

valore centrale, medie algebriche, mediana, moda
(detti anche *misure di intensità, centri . . .*)

- indici di dispersione o variabilità:

intervallo di variazione, varianza, varianza stimata, deviazione standard, deviazione standard stimata

- indici di simmetria o asimmetria: . . .

Valore Centrale

Dato l'insieme di valori $\{x_1, x_2, \dots, x_n\}$, il **valore centrale** considera solo i due valori estremi (non tiene conto di tutti i valori):

$$\frac{x_{\max} + x_{\min}}{2}$$

dove $x_{\max} = \max\{x_1, x_2, \dots, x_n\}$ e $x_{\min} = \min\{x_1, x_2, \dots, x_n\}$.

Esempio: $\{3, 20, 27, 25, 30, 310\}$

$$\frac{x_{\max} + x_{\min}}{2} = \frac{310 + 3}{2} = 156.5$$

Media Aritmetica

Media semplice: dato l'insieme di valori $\{x_1, x_2, \dots, x_n\}$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Media ponderata (dati raggruppati): dato l'insieme di valori $\{x_1, x_2, \dots, x_m\}$ con le rispettive frequenze assolute $\{f_1, f_2, \dots, f_m\}$

$$\bar{x} = \frac{\sum_{i=1}^m f_i x_i}{\sum_{i=1}^m f_i} = \frac{1}{n} \cdot \sum_{i=1}^m f_i x_i = \frac{f_1 x_1 + f_2 x_2 + \dots + f_m x_m}{n}$$

Media Aritmetica – Esercizi

Esercizio 1. Dato l'insieme di valori $\{12, 25, 37, 41, 0, 53\}$, calcolare la media aritmetica. [media aritmetica = 28]

Esercizio 2. Dato l'insieme di valori $\{28, 28, 28, 28, 28, 28\}$, calcolare la media aritmetica.

Esercizio 3. (dati raggruppati) In un campione di 200 persone si sa che 20 pesano 50kg, 30 pesano 55kg, 50 pesano 60kg, 70 pesano 65kg, 20 pesano 75Kg e 10 pesano 80kg. Calcolare il peso medio. [peso medio = 62.5Kg]

Media Aritmetica – Osservazioni

Alcune osservazioni:

- la media può non appartenere all'insieme dei dati
- insiemi di dati diversi possono avere la stessa media
- utilizza tutti i dati
- centro di gravità dei dati
- riduce l'effetto dei dati estremi (*outlier*)

Media Aritmetica – Proprietà

1) Se applico una trasformazione lineare ai dati:

$$y_i = a x_i + b \quad \Rightarrow \quad \bar{y} = a \bar{x} + b$$

2) La somma degli scarti dalla media è nulla:

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

3) La somma dei quadrati degli scarti dalla media è minima:

$$\sum_{i=1}^n (x_i - x)^2 \text{ assume il valore minimo per } x = \bar{x}$$

Media Aritmetica – Proprietà

- La somma degli scarti dalla media è nulla:

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - \sum_{i=1}^n \bar{x} = \sum_{i=1}^n x_i - n\bar{x} = \sum_{i=1}^n x_i - n \cdot \frac{1}{n} \cdot \sum_{i=1}^n x_i = 0$$

- La somma dei quadrati degli scarti dalla media è minima:

poniamo $g(x) = \sum_{i=1}^n (x_i - x)^2$. Abbiamo che

$$g(x) = \sum_{i=1}^n (x_i)^2 - 2 \sum_{i=1}^n x_i x + \sum_{i=1}^n x^2 = \sum_{i=1}^n (x_i)^2 - 2n\bar{x}x + nx^2$$

Quindi, g è un polinomio di secondo grado in x .

Pertanto, assume il suo valore minimo in $x = -\frac{-2n\bar{x}}{2n} = \bar{x}$.

Media Geometrica

Media semplice: dato l'insieme di valori $\{x_1, x_2, \dots, x_n\}$ con la condizione che siano *tutti positivi*

$$x_g = \sqrt[n]{\prod_{i=1}^n x_i} = \sqrt[n]{x_1 \cdot x_2 \cdots x_n} \quad \Rightarrow \quad \log x_g = \frac{1}{n} \cdot \sum_{i=1}^n \log x_i$$

Media ponderata: dato l'insieme di valori $\{x_1, x_2, \dots, x_m\}$, tutti positivi, con le rispettive frequenze assolute $\{f_1, f_2, \dots, f_m\}$

$$x_g = \sqrt[n]{\prod_{i=1}^m (x_i)^{f_i}} \quad \Rightarrow \quad \log x_g = \frac{1}{n} \cdot \sum_{i=1}^m f_i \log x_i$$

Mediana

Dato l'insieme di valori ordinati $x_1 \leq x_2 \leq \dots \leq x_{n-1} \leq x_n$, si chiama **mediana** (o valore mediano) il valore M_e che occupa la posizione centrale:

- se n è dispari, c'è un unico termine mediano di posto $\frac{n+1}{2}$

$$M_e = x_{\frac{n+1}{2}}$$

- se n è pari ci sono due termini mediani di posti $\frac{n}{2}$ e $\frac{n}{2} + 1$

$$M_e = \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1})$$

Utilizza tutti i valori ma si basa soltanto sull'ordinamento degli stessi.

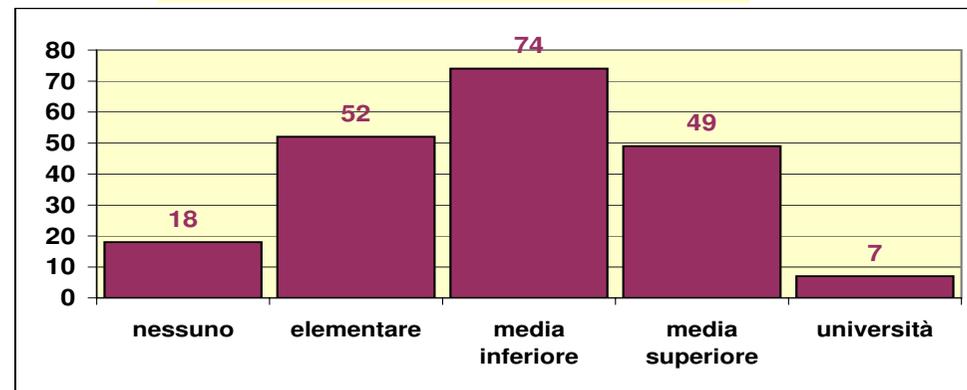
Esempio 1. $\{503, 25, 0, 81, 13\} \Rightarrow M_e = 25$

Esempio 2. $\{327, 2, 93, 1, 503, 81\} \Rightarrow M_e = 87$

Moda

Moda: valore (o classe) al quale è associata la frequenza più alta

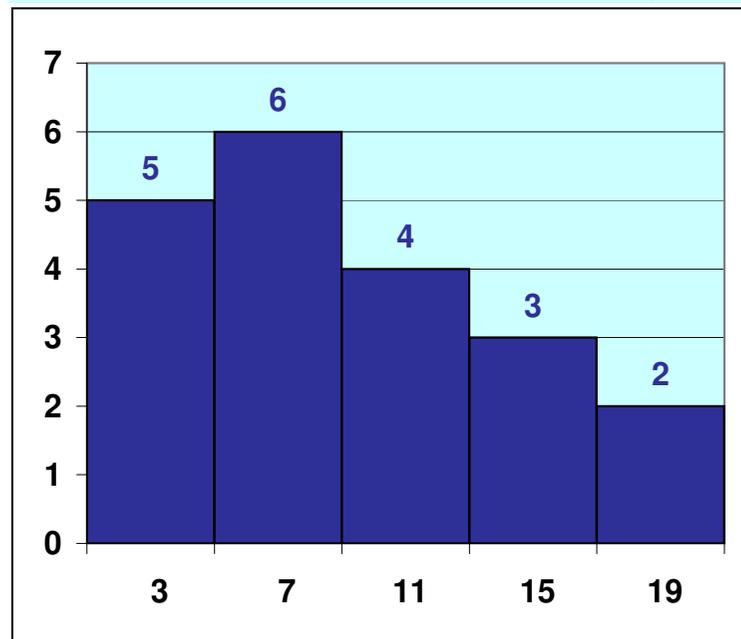
titolo di studio	fi	fi / n
nessuno	18	0,09
elementare	52	0,260
media inferiore	74	0,370
media superiore	49	0,245
università	7	0,035
	200	1,000



Si può applicare anche a dati qualitativi espressi su scala nominale.

Esempio: Media, Mediana, Moda

classe	r_i	f_i	f_i / n
1-5	3	5	0,25
5-9	7	6	0,300
9-13	11	4	0,200
13-17	15	3	0,150
17-21	19	2	0,100
		20	1,000



- **Media:** $\bar{x} = 9.2$
si calcola come media ponderata
- **Mediana:** $M_e = 7$
è la media del decimo e dell'undicesimo termine che hanno entrambi valore 7
- **Moda:** è la classe 5 – 9 o il suo rappresentante $r_2 = 7$, corrispondenti a $f_2 = 6$
- *moda < mediana < media*
distribuzione obliqua a destra

Esercizi

Esercizio 1. Vengono intervistati 50 capofamiglia, ponendo la seguente domanda: *Quanti figli ci sono nella sua famiglia?*

numero figli	0	1	2	3	4	5	6
f_{assoluta}	6	12	16	9	4	1	2

Calcolare: frequenze relative, frequenze cumulate, valore centrale, media, mediana.

Esercizi

Soluzione:

numero figli	0	1	2	3	4	5	6
f_{assoluta}	6	12	16	9	4	1	2
f_{relativa}	0.12	0.24	0.32	0.18	0.08	0.02	0.04
F_{cumulata}	6	18	34	43	47	48	50

$$\text{valore centrale} = \frac{0 + 6}{2} = 3.00$$

$$\text{mediana} = \frac{x_{25} + x_{26}}{2} = 2.00$$

$$\text{media} = \frac{1}{50} (0 + 12 + 32 + 27 + 16 + 5 + 12) = 2.08$$

Esercizi

Esercizio 2. Vengono intervistati i figli delle stesse 50 famiglie, ponendo la seguente domanda: *Quanti figli siete in famiglia?*

Costruire la tabella delle frequenze e calcolare mediana, media e valore centrale.

Esercizi

Soluzione:

numero figli	1	2	3	4	5	6
f_{assoluta}	12	32	27	16	5	12
F_{cumulata}	12	44	71	87	92	104

valore centrale = 3.50

$$\text{mediana} = \frac{x_{52} + x_{53}}{2} = 3.00$$

$$\text{media} = 3.06$$

- è cambiata la popolazione
- le famiglie più numerose pesano di più
- si perde l'informazione sulle famiglie senza figli

Statistica – Un Esempio

Un'indagine sul peso, su un campione di $n = 100$ studenti, ha prodotto il seguente risultato. I pesi p sono espressi in Kg e sono stati raggruppati in cinque *classi* di peso.

classe (peso in Kg)	f_i	f_i/n
$60 \leq p \leq 62$	5	0.05
$63 \leq p \leq 65$	18	0.18
$66 \leq p \leq 68$	42	0.42
$69 \leq p \leq 71$	27	0.27
$72 \leq p \leq 74$	8	0.08
	100	1.00

Sono riportate le frequenze assolute f_i (numero di individui appartenenti alla classe di peso i -sima) e le frequenze relative f_i/n .

Le classi sono di uguale ampiezza, ma non sono contigue

Statistica – Un Esempio

Estendiamo i confini di ciascuna classe in modo simmetrico di 0.5 Kg. La popolazione non è cambiata e possiamo rappresentare i dati, in classi contigue, come segue:

classe (peso in Kg)	r_i	f_i	f_i/n
$59.5 \leq p < 62.5$	61	5	0.05
$62.5 \leq p < 65.5$	64	18	0.18
$65.5 \leq p < 68.5$	67	42	0.42
$68.8 \leq p < 71.5$	70	27	0.27
$71.5 \leq p < 74.5$	73	8	0.08
		100	1.00

Supponendo che gli individui di una classe siano distribuiti *uniformemente* al suo interno, è naturale associare a ciascuna classe, come *rappresentante*, il valore centrale r_i della classe stessa.

Calcolo della Media

Come si può calcolare la media dei dati conoscendo solo un'informazione parziale (*per classi*) sulle frequenze?

Occorre formulare un'ipotesi su come i dati si distribuiscono all'*interno* di ogni classe. In assenza di ulteriori informazioni, è ragionevole congetturare che gli elementi appartenenti ad una classe si distribuiscano *uniformemente* al suo interno.

È naturale associare ad ogni classe un *rappresentante*: il valore centrale della classe.

r_i	61	64	67	70	73
f_i	5	18	42	27	8

Ai fini del calcolo della media si utilizzano solo i rappresentanti r_i :

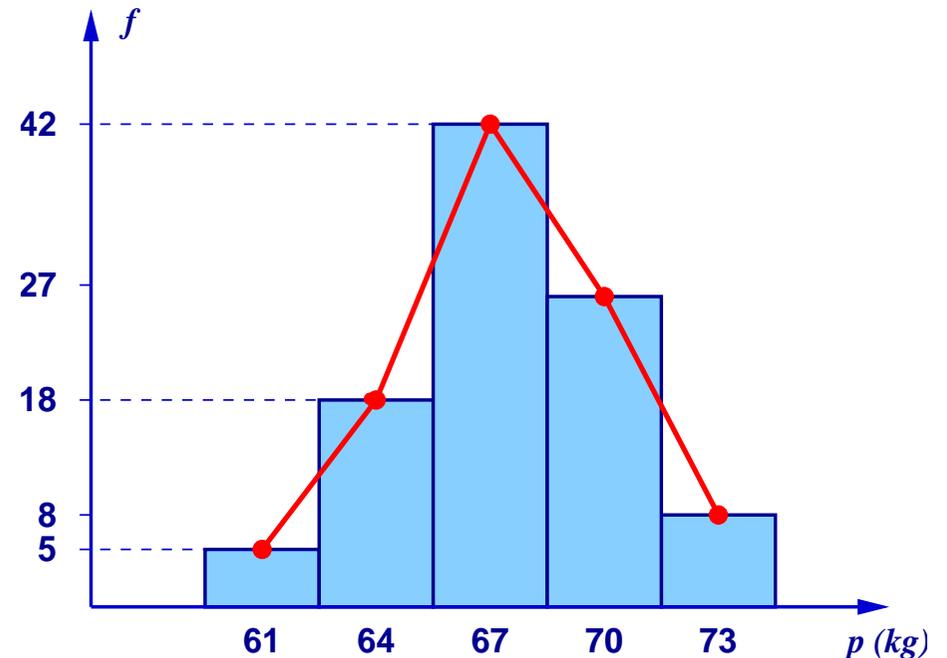
$$\bar{p} = \frac{5 \cdot 61 + 18 \cdot 64 + 42 \cdot 67 + 27 \cdot 70 + 8 \cdot 73}{100} = 67.45 \text{ Kg}$$

Poligono di Frequenza

Distribuzione delle frequenze

classe	r_i	f_i
$59.5 \leq p < 62.5$	61	5
$62.5 \leq p < 65.5$	64	18
$65.5 \leq p < 68.5$	67	42
$68.8 \leq p < 71.5$	70	27
$71.5 \leq p < 74.5$	73	8

- possiamo rappresentare in modo efficace le frequenze delle classi del campione mediante un **istogramma** (dove le aree dei rettangoli sono proporzionali alle frequenze della classe)
- unendo i punti medi • dei lati superiori dei rettangoli, si ottiene il cosiddetto **poligono di frequenza**



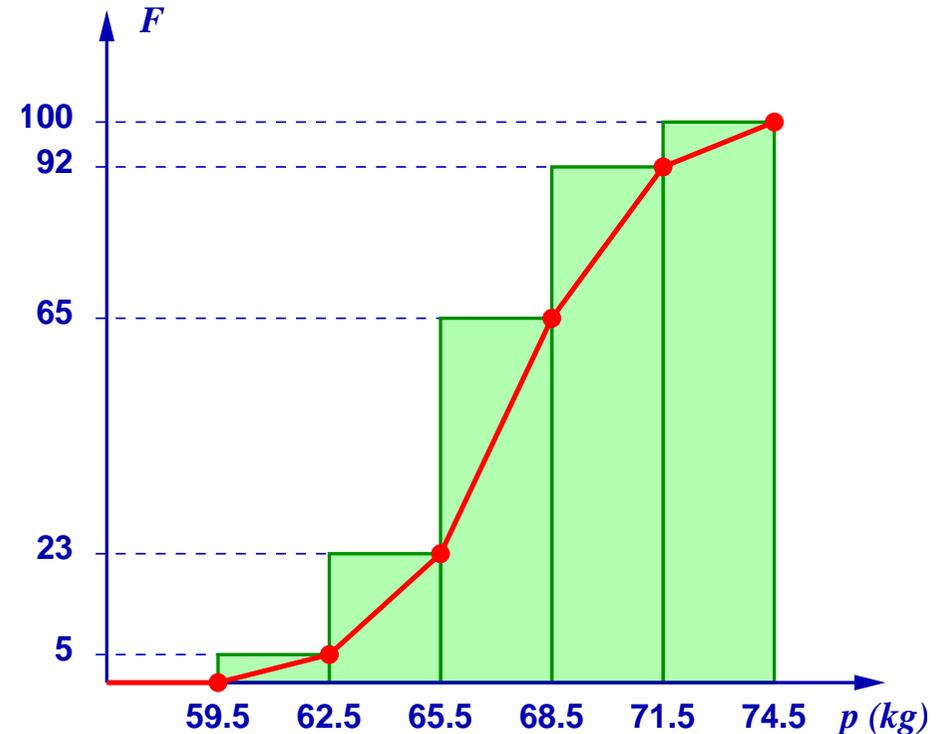
Ipotesi: classi equispaziate – distribuzione uniforme all'interno della classe

Ogiva di Frequenza

Distribuzione delle frequenze

classe	r_i	f_i	F_i
$p < 59.5$	-	0	0
$59.5 \leq p < 62.5$	61	5	5
$62.5 \leq p < 65.5$	64	18	23
$65.5 \leq p < 68.5$	67	42	65
$68.8 \leq p < 71.5$	70	27	92
$71.5 \leq p < 74.5$	73	8	100

- calcoliamo le *frequenze cumulate* F_i (F_i rappresenta il numero dei dati, che sono minori del secondo estremo della i -sima classe)
- costruiamo il **diagramma cumulativo** delle frequenze
- unendo i punti ●, si ottiene la cosiddetta **ogiva di frequenza**



Calcolo della Mediana: Primo Metodo

Calcolo della mediana M_e

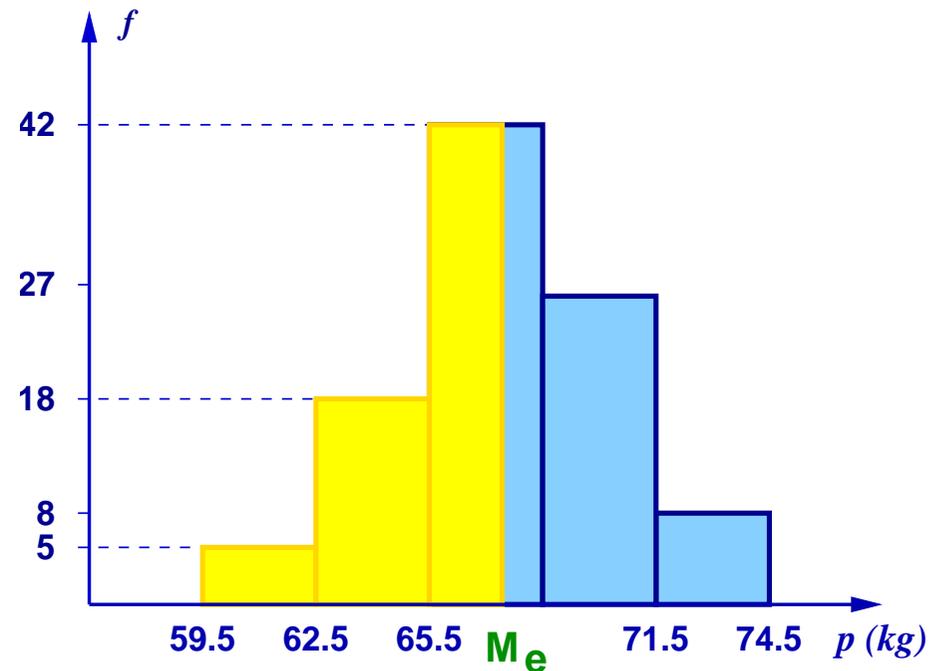
Trovare il punto M_e tale che l'area in giallo sia il 50% dell'area totale sottesa dall'istogramma delle frequenze

area totale istogramma = 300

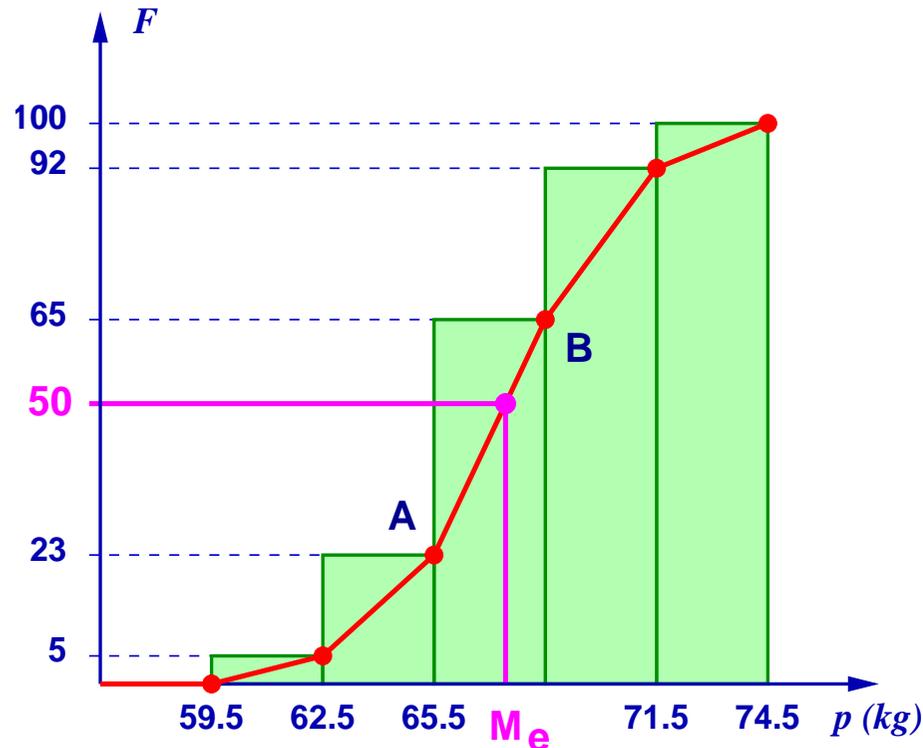
$$3 \cdot 5 + 3 \cdot 18 + (M_e - 65.5) \cdot 42 = 150$$

$$\Rightarrow M_e = \frac{81}{42} + 65.5 \simeq 67.43 \text{ Kg}$$

NOTA: ricordiamo che le aree sono proporzionali alle frequenze.



Calcolo della Mediana: Secondo Metodo



Calcolo della mediana M_e

Trovare il punto di intersezione della retta $F = 50$ con l'ogiva di frequenza.

Significa trovare l'intersezione con la retta passante per i punti:

$A = (65.5, 23)$ e $B = (68.5, 65)$

$$\begin{cases} F = 50 \\ F = 23 + \frac{42}{3} \cdot (p - 65.5) \end{cases}$$

$$\Rightarrow p = 65.5 + \frac{3}{42} \cdot 27 \simeq 67.43$$

Indici di Dispersione

Si cercano indici di dispersione che:

- utilizzino tutti i dati $\{x_1, x_2, \dots, x_n\}$
- 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 $\{x_1, x_2, \dots, x_n\}$

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

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

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

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

$$y_i = ax_i + b \quad \Rightarrow \quad s_y^2 = a^2 s_x^2, \quad s_y = |a| s_x$$

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:

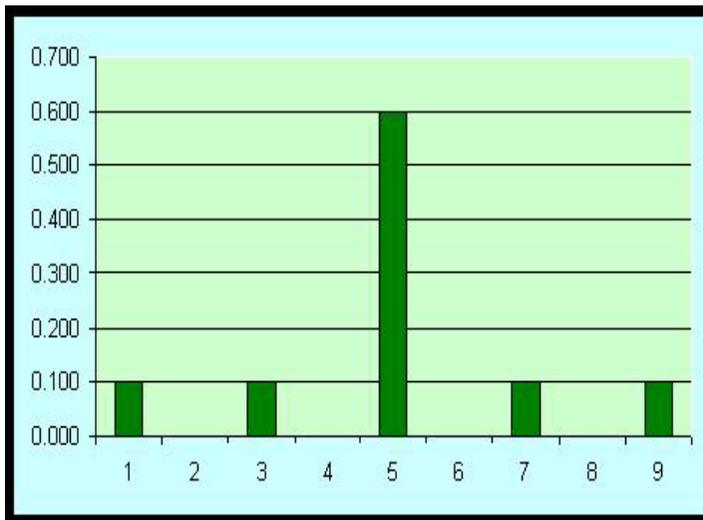
Varianza campionaria (*stimata*):

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

Deviazione standard campionaria (*stimata*):

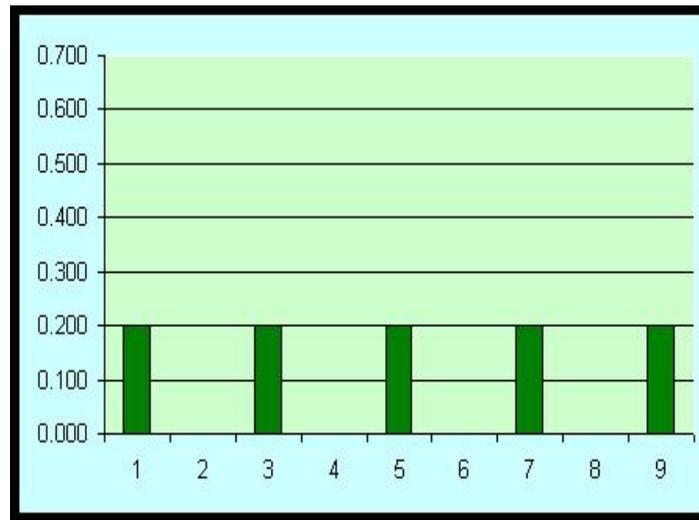
$$s = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2}$$

Esempio Riassuntivo



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	

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

media	5.00
mediana	5.00
varianza	4.00
varianza stimata	4.44
deviazione standard	2.00
deviazione standard stimata	2.11

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

media	5.00
mediana	5.00
varianza	8.00
varianza stimata	8.89
deviazione standard	2.83
deviazione standard stimata	2.98

Esercizi

Esercizio 1. 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*)

Calcolare il peso medio e lo scarto quadratico medio.

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:

$$\bar{p} = \frac{1}{100} (20 \cdot 52.5 + 15 \cdot 57.5 + 18 \cdot 62.5 + 22 \cdot 67.5 + 18 \cdot 72.5 + 7 \cdot 77.5) = 63.7 \text{ Kg}$$

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

$$s^2 = \frac{1}{100} (20 \cdot 125.44 + 15 \cdot 38.44 + 18 \cdot 1.44 + 22 \cdot 14.44 + 18 \cdot 77.44 + 7 \cdot 190.44) \simeq 61.56 \text{ Kg}^2$$

$$s \simeq 7.85 \text{ Kg}$$

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 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:

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

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

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

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

$q_3 = \frac{x_{33} + x_{34}}{2} = 172$ 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 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

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

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

$$s \simeq 3.69$$

$$\text{moda} = 10.0$$

$$\text{mediana} = 8.5$$

Esercizi

Esercizio 3. 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 (voto in centesimi)	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

Calcolare media e varianza. Usando l'istogramma delle frequenze o l'ogiva di frequenza, calcolare la mediana. Calcolare i quartili dall'ogiva di frequenza. Qual è il voto minimo che bisogna aver preso per non far parte del 10% degli studenti peggiori?

Esercizi

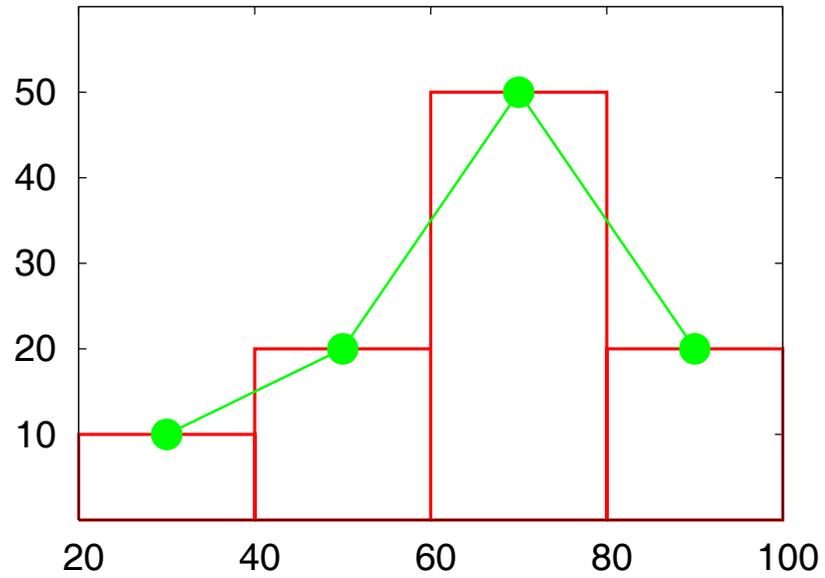
Soluzione: le classi sono di uguale ampiezza e contigue. Nell'ipotesi di *distribuzione uniforme*, è naturale associare a ciascuna classe, come *rappresentante*, il valore centrale r_i della classe stessa.

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

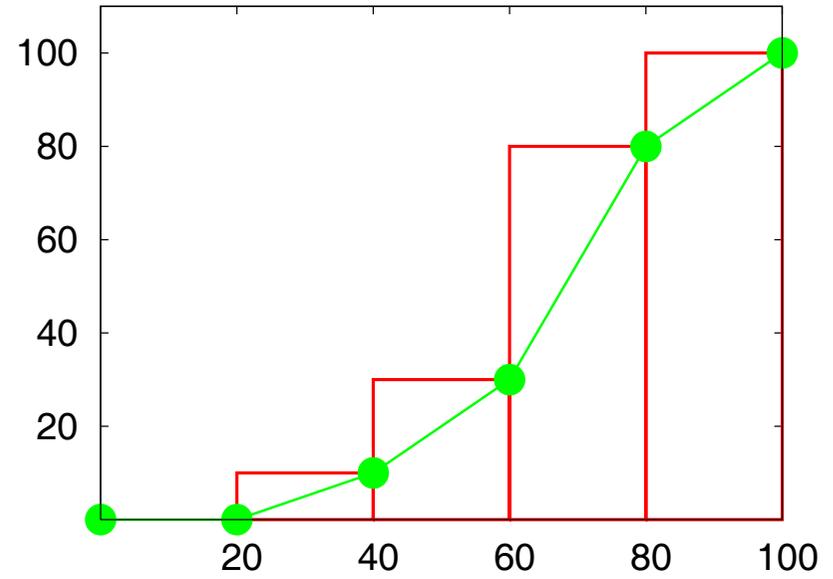
$$\text{media} = \frac{1}{100} \cdot (10 \cdot 30 + 20 \cdot 50 + 50 \cdot 70 + 20 \cdot 90) = 66$$

$$\text{varianza} = \frac{1}{100} \cdot (10 \cdot 36^2 + 20 \cdot 16^2 + 50 \cdot 4^2 + 20 \cdot 24^2) = 304$$

Esercizi



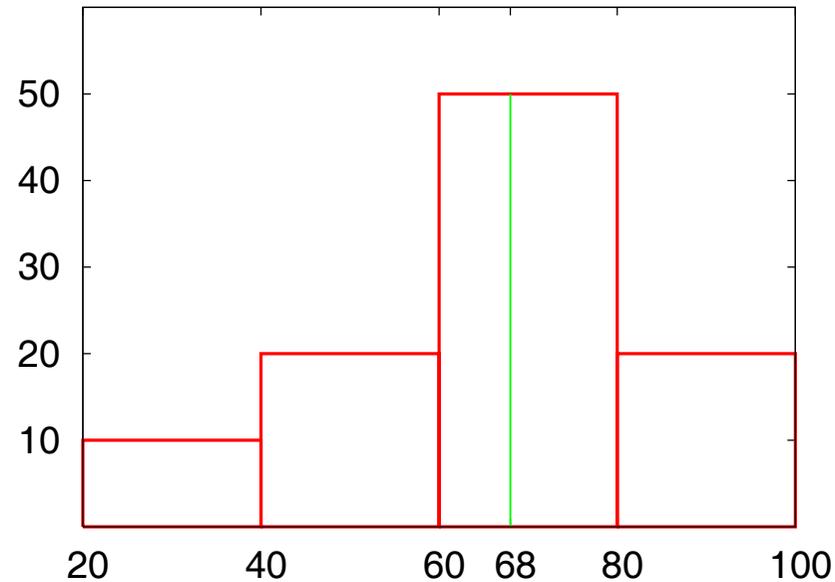
Poligono delle frequenze



Ogiva di frequenza

Esercizi

Calcolo della mediana: con l'istogramma delle frequenze



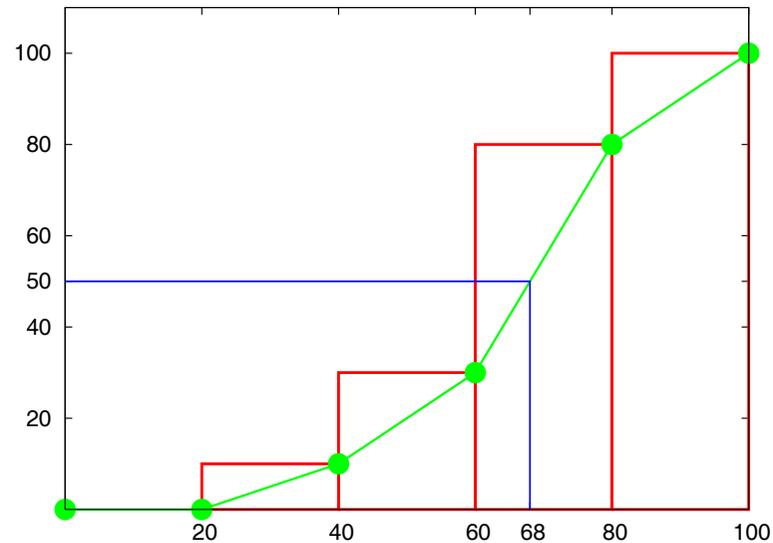
$$\text{area totale} = 20 \cdot (10 + 20 + 50 + 20) = 2000$$

Cerchiamo il valore $x = M_e$ tale che

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

Esercizi

Calcolo della mediana: con l'ogiva di frequenza



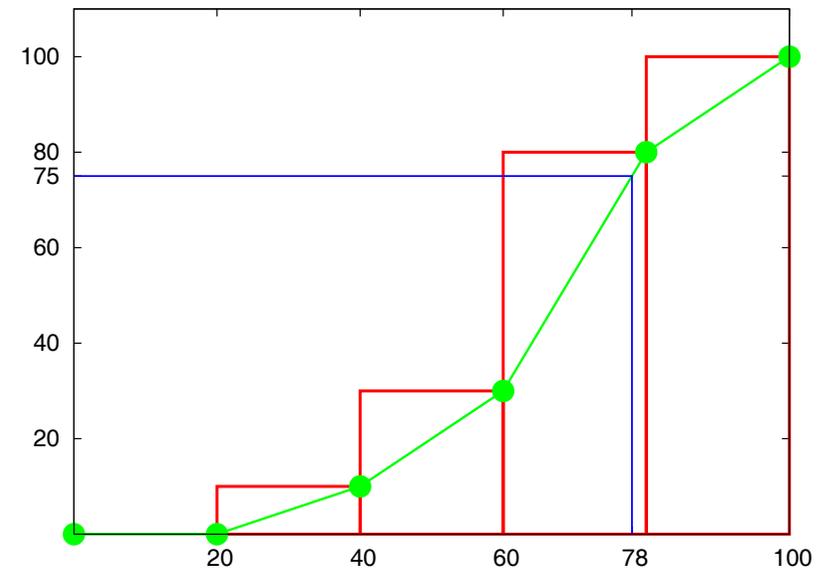
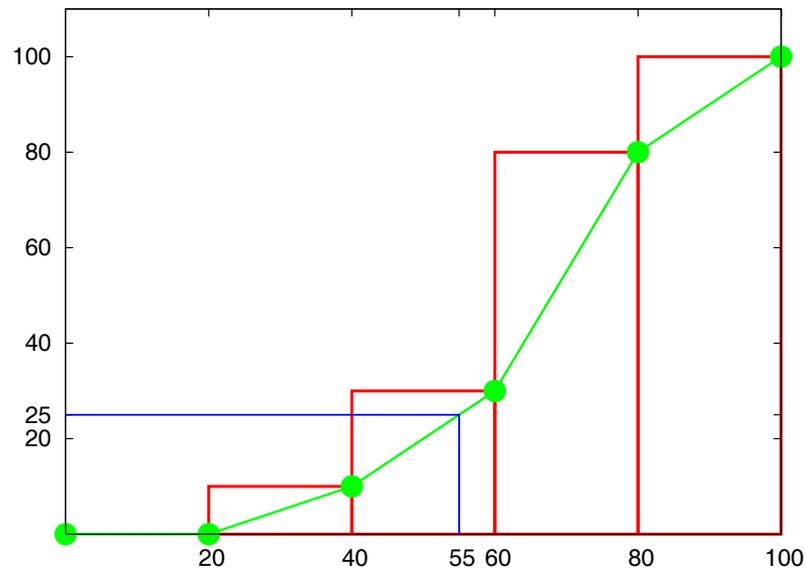
Si considera l'interpolazione lineare sui punti $A = (60, 30)$ e $B = (80, 80)$

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

Esercizi

Calcolare i quartili dall'ogiva di frequenza.

$$q_1 = 55, q_3 = 78$$



Esercizi

Qual è il voto minimo che bisogna aver preso per non far parte del 10% degli studenti peggiori? risposta: 40

