

## Statistica - Un Esempio

---



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

<i>classe</i>	$f_i$	$f_i/n$
$60 \leq p \leq 62$	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 la frequenze relative  $f_i/n$ .

**Le classi sono di uguale ampiezza, ma non sono contigue**

## Statistica - Un Esempio



Estendiamo i confini di ciascuna classi in modo simmetrico di  $0.5 \text{ kg}$ . La popolazione non è cambiata e possiamo rappresentare i dati, in classi contigue, come segue:

<i>classe</i>	$r_i$	$f_i$	$f_i/n$
$59.5 \leq p < 62.5$	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	8	42	27	8

---

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

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

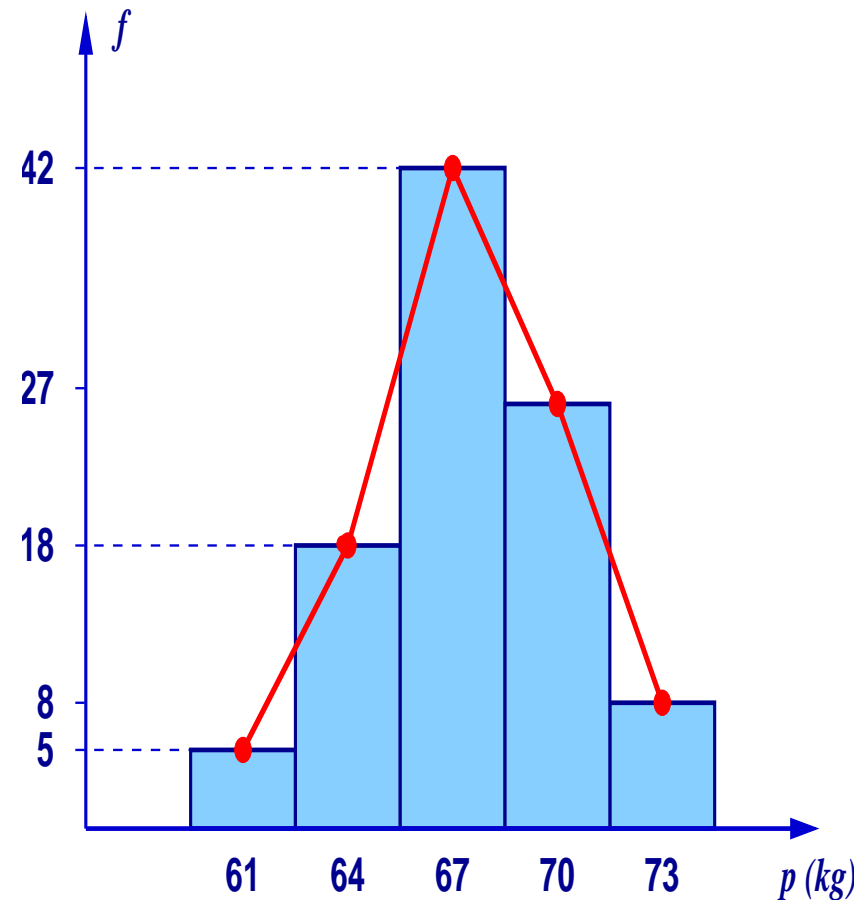
# Poligono delle Frequenze



## DISTRIBUZIONE FREQUENZE

<i>classe</i>	<i>r<sub>i</sub></i>	<i>f<sub>i</sub></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

- posso 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, ottengo il cosiddetto **poligono delle frequenze**



**IPOSTESI: classi equispaziate – distribuzione uniforme all'interno della classe**

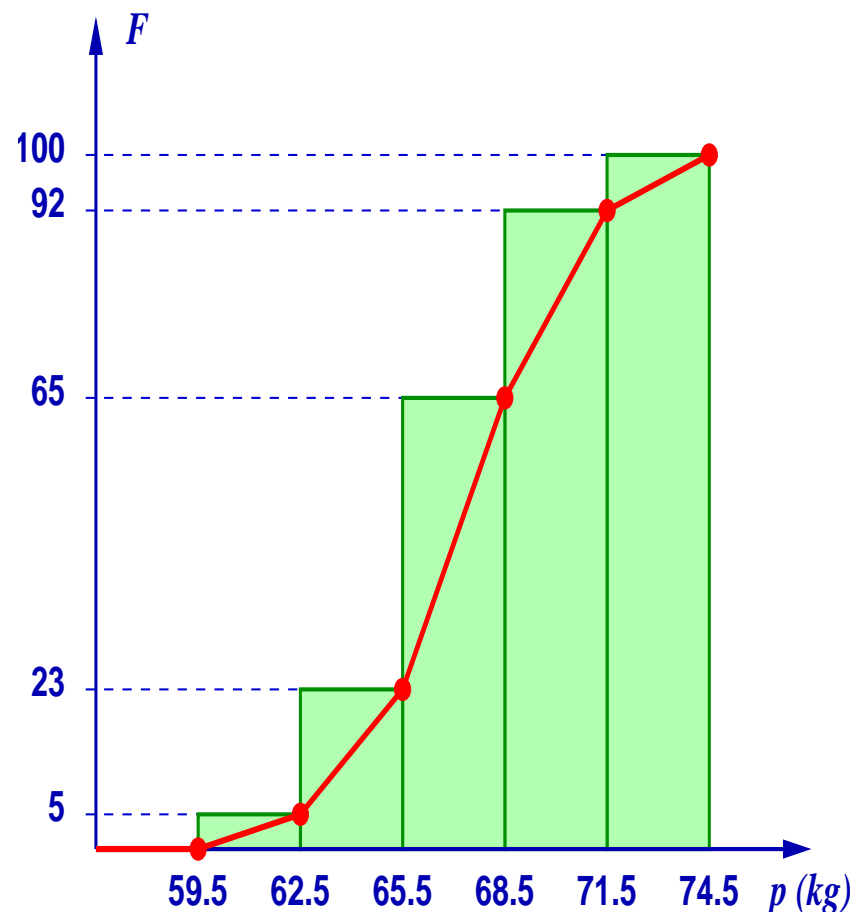
# Ogiva di Frequenza



## DISTRIBUZIONE FREQUENZE

<i>classe</i>	$r_i$	$f_i$	$F_i$
$p < 59.5$	-	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

- calcolo le *frequenze cumulate*  $F_i$ , ( $F_i$  rappresenta il numero dei dati, che sono minori del secondo estremo della  $i$ -sima classe)
- costruisco il **diagramma cumulativo** delle frequenze
- unendo i punti ●, ottengo la cosiddetta **ogiva di frequenze**



# Calcolo della Mediana - 1



## CALCOLO DELLA MEDIANA $M_e$ :

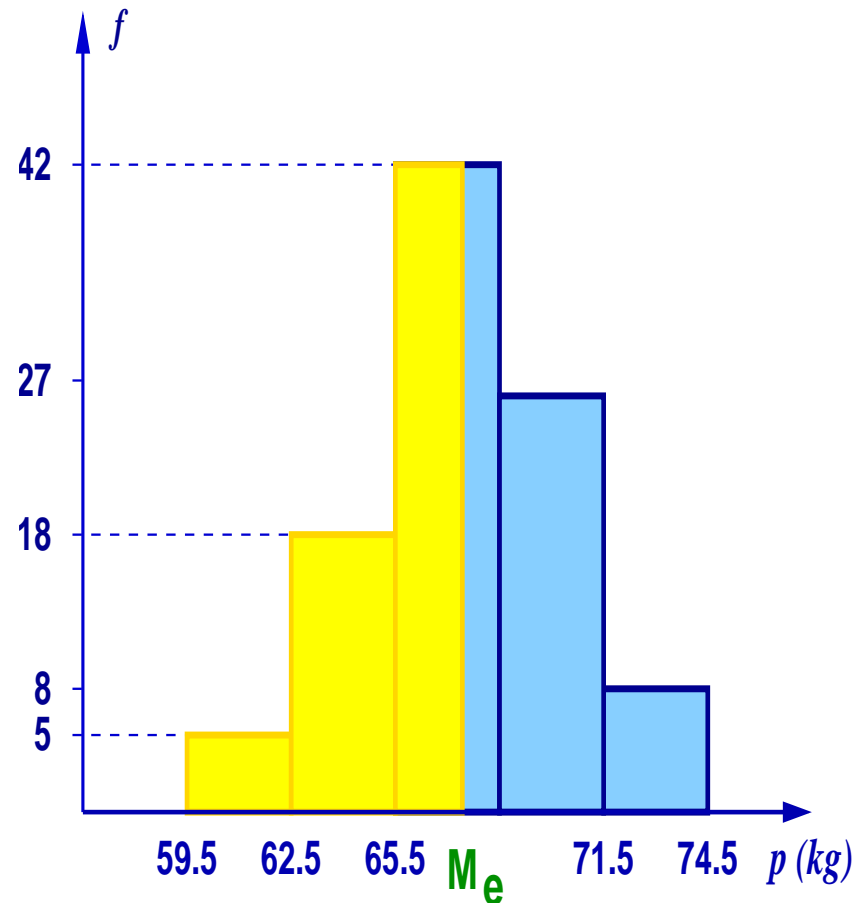
Trovare il punto  $M_e$  tale che l'area in giallo sia il 50% dell'area totale sottesa dall'istogramma delle frequenze

$$\text{area totale istogramma} = 300$$

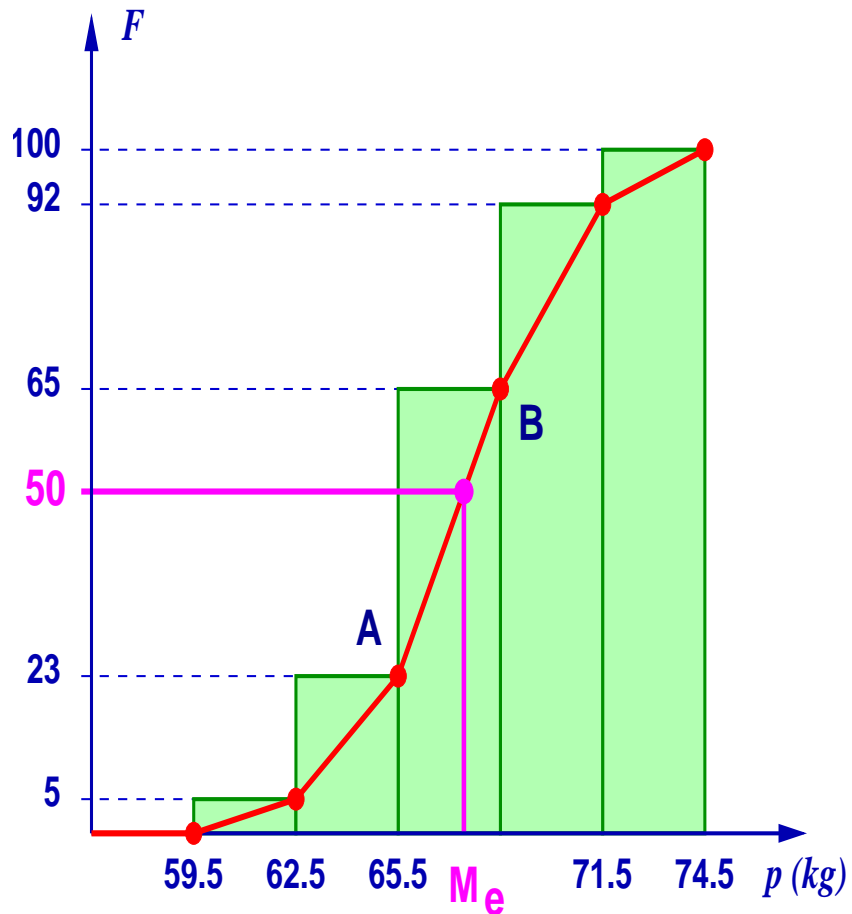
$$3 \cdot 5 + 3 \cdot 18 + (M_e - 65.5) \cdot 42 = 150$$

$$\Rightarrow M_e = \frac{81}{42} + 65.5 \approx 67.43 \text{ kg}$$

**NOTA:** ricordiamo che le aree sono proporzionali alle frequenze.



## Calcolo della Mediana - 2



### 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:

$$A = (65.5, 23) \text{ e } B = (68.5, 65)$$

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

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

## Esercizio 1



**ESERCIZIO** – Vengono intervistati 50 capofamiglia, ponendo la seguente domanda: *Quanti figli ci sono nella sua famiglia ?*

<i>numero figli</i>	0	1	2	3	4	5	6
<i>f<sub>assoluta</sub></i>	6	12	16	9	4	1	2

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



## Esercizio 1



### SOLUZIONE

<i>numero figli</i>	0	1	2	3	4	5	6
<i>f<sub>assoluta</sub></i>	6	12	16	9	4	1	2
<i>f<sub>relativa</sub></i>	0.12	0.24	0.32	0.18	0.08	0.02	0.10
<i>F<sub>cumulata</sub></i>	6	18	34	43	47	48	<b>50</b>

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

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

$$\text{media} = \frac{1}{50} (0 + 12 + 32 + 27 + 16 + 5 + 12) = 2.0$$

## Esercizio 2

---



**ESERCIZIO** – Vengono intervistati i figli delle stesse 50 famiglie, ponendo la seguente domanda: *Quanti fratelli siete in famiglia ?*

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

## Esercizio 2



### SOLUZIONE

<i>numero fratelli</i>	1	2	3	4	5	6
<i>f<sub>assoluta</sub></i>	12	32	27	16	5	12
<i>F<sub>cumulata</sub></i>	12	44	71	87	92	<b>104</b>

*valore centrale* = 3.00

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

$$\text{media} = = 3.06$$

1. è cambiata la popolazione
2. le famiglie più numerose pesano di più
3. si perde l'informazione sulle famiglie senza figli

## Esercizio 3



**ESERCIZIO** – Uno stato è suddiviso in due aree geografiche: Nord e Sud. L'età media al Nord è di 42.5 anni, l'età media al Sud è di 48.5 anni. Sapendo che gli abitanti del Nord sono 45 milioni e quelli del Sud sono 55 milioni, è possibile stimare l'età media dell'intera popolazione ?

**SOLUZIONE** – La risposta *SI* (le informazioni a disposizione sono sufficienti) e si procede come segue

	<i>età media</i>	<i>f<sub>ass</sub></i> ( <i>numero abitanti</i> )	<i>f<sub>rel</sub></i>
Nord	42.5	45 milioni	0.45
Sud	48.5	55 milioni	0.55

$$\text{età media} = 0.45 \cdot 42.5 + 0.55 \cdot 48.5 = 45.8 \quad (45 \text{ anni e } 10 \text{ mesi})$$