

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	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 0.5kg. La popolazione non è cambiata e possiamo rappresentare i dati, in classi contigue, come segue:

classe	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 :

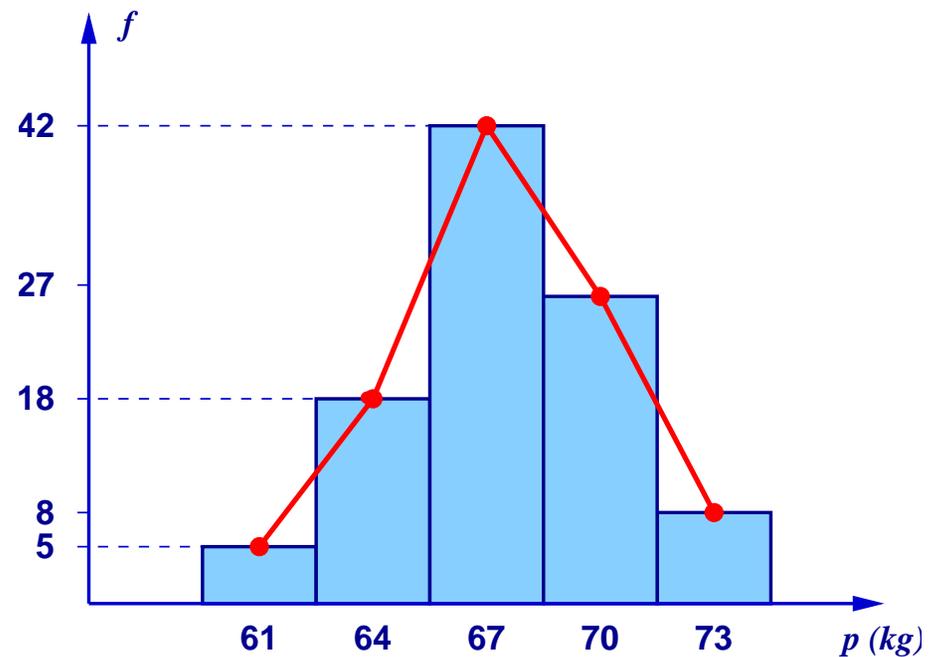
$$\bar{p} = \frac{5 \cdot 61 + 18 \cdot 64 + 42 \cdot 67 + 27 \cdot 70 + 8 \cdot 73}{100} = 67.45\text{kg}$$

Poligono di Frequenza

DISTRIBUZIONE 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**



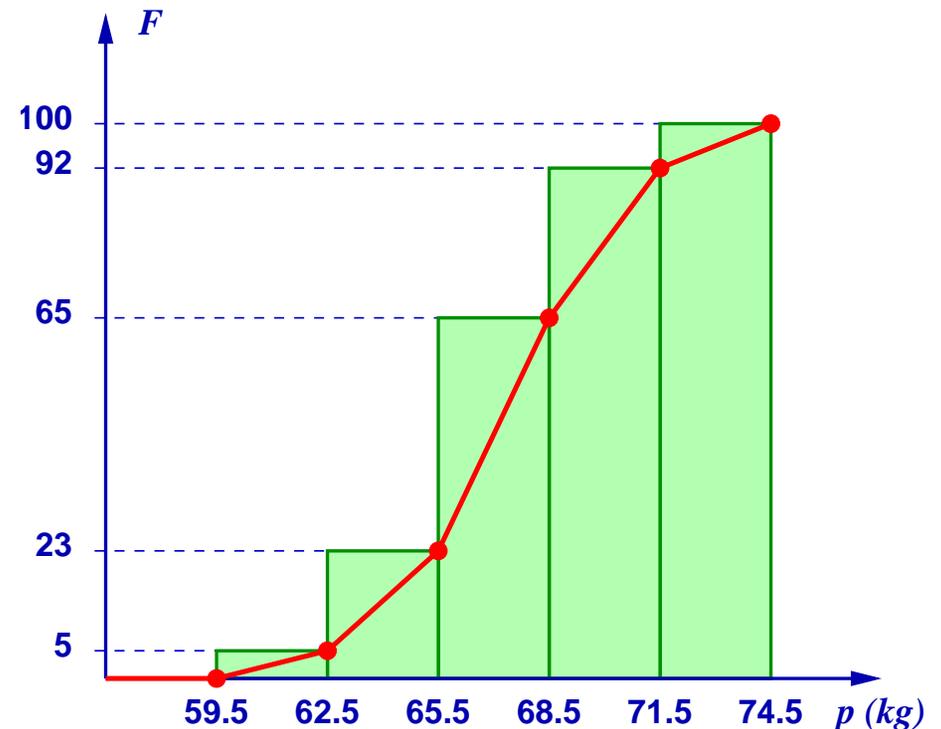
IPOSTESI: classi equispaziate – distribuzione uniforme all'interno della classe

Ogiva di Frequenza

DISTRIBUZIONE 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 – 1

CALCOLO DELLA MEDIANA M_e

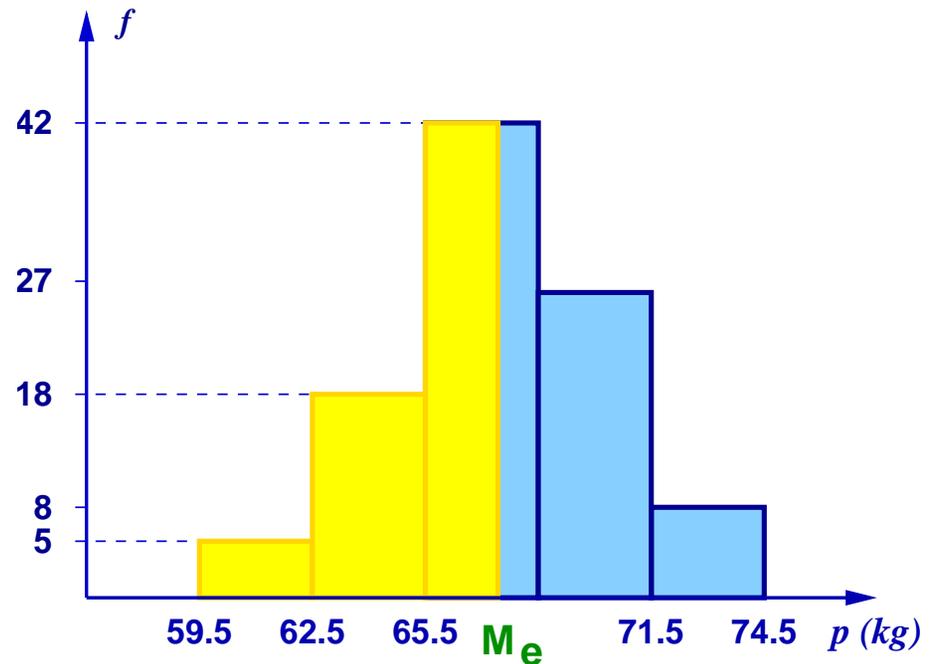
Trovare il punto M_e tale che l'area in giallo sia il 50% dell'area totale sottesa dall'istogramma delle frequenze

area totale istogramma = 300

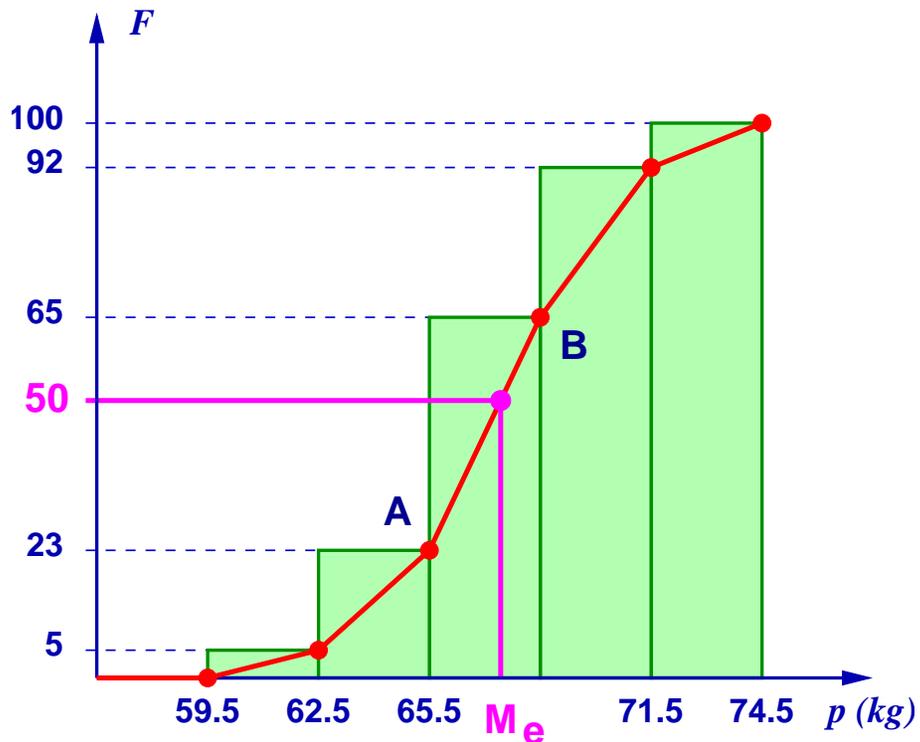
$$3 \cdot 5 + 3 \cdot 18 + (M_e - 65.5) \cdot 42 = 150$$

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

NOTA: ricordiamo che le aree sono proporzionali alle frequenze.



Calcolo della Mediana – 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:

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

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

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

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

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

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

$$\text{media} = \frac{1}{50} (0 + 12 + 32 + 27 + 16 + 5 + 12) = 2.00$$

Esercizi

Esercizio 2. 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.

Esercizi

Soluzione:

numero fratelli	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

Esercizi

Esercizio 3. 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 è sì (le informazioni a disposizione sono sufficienti). Si procede come segue

	età media	f_{ass} (numero abitanti)	f_{rel}
Nord	42.5	45 milioni	0.45
Sud	48.5	55 milioni	0.55

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