

Corso di Statistica Psicometrica

Canale 1 - Matricole Dispari

CdS triennale in
Scienze e Tecniche
Psicologiche

Domenico Vistocco

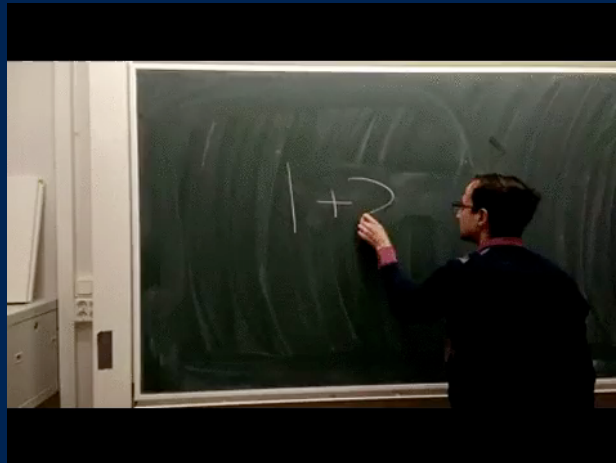
Dipartimento di Scienze Politiche
Università degli Studi di Napoli Federico II



Nella
puntata
precedente...



- Tipologia di variabili
- Tipologia di tabelle
- Operatore sommatoria e principali proprietà
- Cenni ai criteri per la costruzione di tabelle di frequenza per classi di modalità
- Ipotesi di equidistribuzione all'interno delle classi ed uso dei centri come elementi rappresentativi delle classi
- Un esempio di "tecnicismo": le differenti formule per il calcolo della media aritmetica



Si riparte?

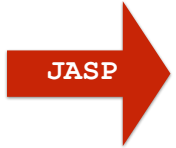
Software



<https://www.jamovi.org/>

Disponibile solo per PC a 64 bit (da un poco di anni è lo standard sul mercato)

Chi ha problemi con l'installazione (PC a 32 bit) può utilizzare questa alternativa:



<https://jasp-stats.org/>

Nella sezione Download è disponibile una versione a 32 bit

[Consulta i tutorial per ulteriori dettagli](#)

PRINCIPALI RAPPRESENTAZIONI GRAFICHE
(vedi il file html con gli esempi)

Variabili qualitative

Diagramma a barre, diagramma a torta, diagramma di Pareto

Variabili quantitative

Diagramma ad aste (a bastoncini), dotplot, istogramma,
poligono di frequenza, funzione di ripartizione empirica

E tu che frequenza sei?

Frequenze e dintorni

FREQUENZE SEMPLICI

$$n_i \quad \left(\sum_{i=1}^k n_i = n \right)$$

$$f_i = \frac{n_i}{n} \quad \left(\sum_{i=1}^k f_i = 1 \right)$$

$$p_i = \frac{n_i}{n} \times 100 = f_i \times 100$$
$$\left(\sum_{i=1}^k p_i = 100 \right)$$

FREQUENZE CUMULATE

$$N_i = \sum_{j=1}^k n_j \quad (N_1 = n_1, N_k = n)$$

$$F_i = \sum_{j=1}^k f_j \quad (F_1 = f_1, F_k = 1)$$

$$P_i = \sum_{j=1}^k p_j \quad (P_1 = p_1, P_k = 100)$$

Cumuliamo...

Per variabili qualitative ordinali
e variabili quantitative

VOTO	Frequenza assoluta	Frequenza cumulata
18	5	5
23	21	26
25	10	36
27	14	50
28	18	68
30	11	79
	79	

Cumuliamo...

Colore capelli	Frequenza assoluta	Frequenza percentuale
Biondi	15	19.0
Castani	55	69.6
Neri	4	5.1
Rossi	5	6.3
	79	100

Colore capelli	Frequenza assoluta	Frequenza percentuale	Frequenza percentuale cumulata
Castani	55	69.6	69.6
Biondi	15	19.0	88.6
Rossi	5	6.3	94.9
Neri	4	5.1	100
	79	100	

Eccezione variabili qualitative nominali (utile per il diagramma di Pareto)

(al limite delle) frequenze

PROPRIETA' DELLA FUNZIONE DI RIPARTIZIONE (empirica)

$$\lim_{x \rightarrow -\infty} N(x) = 0$$

$$\lim_{x \rightarrow +\infty} N(x) = n$$

$$\lim_{x \rightarrow -\infty} F(x) = 0$$

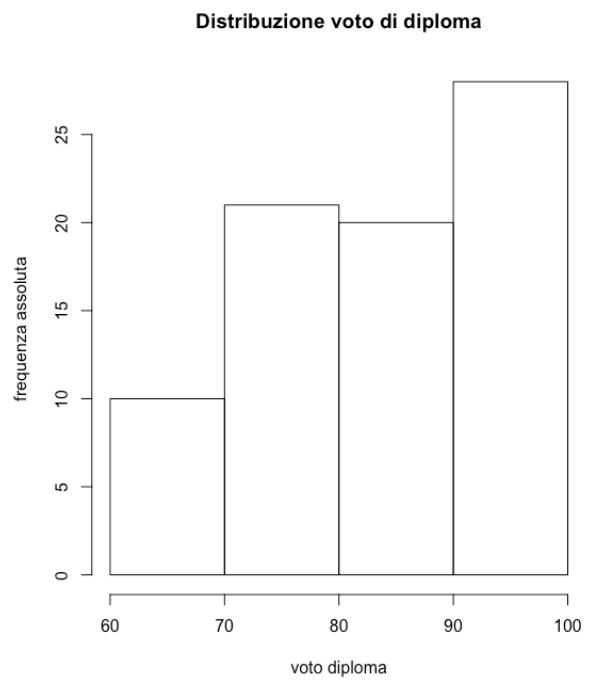
$$\lim_{x \rightarrow +\infty} F(x) = 1$$

$$\lim_{x \rightarrow -\infty} P(x) = 0$$

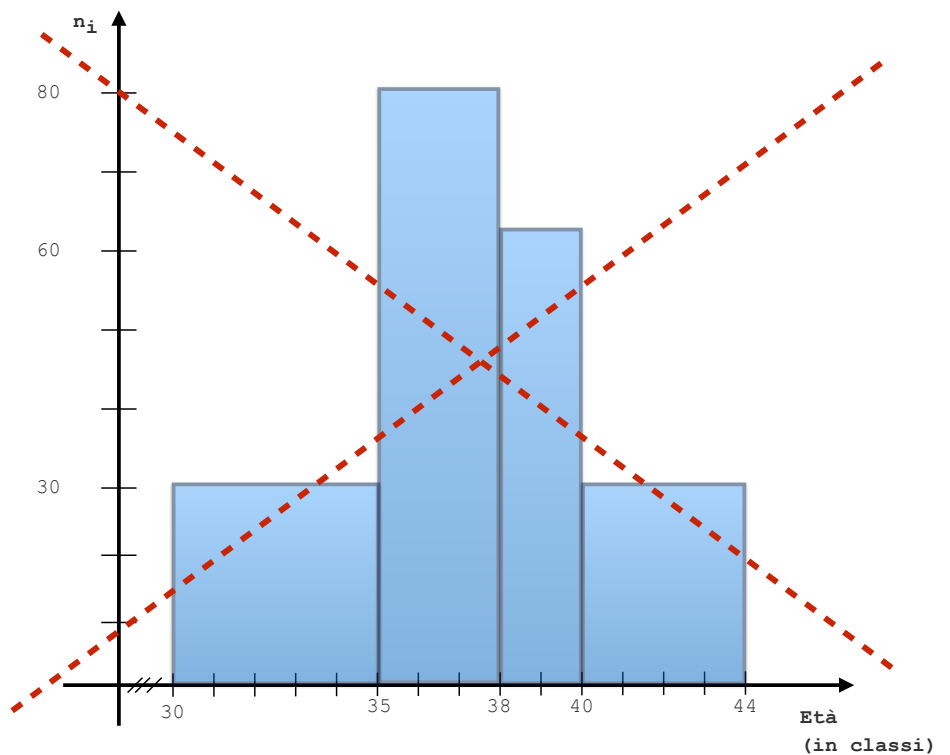
$$\lim_{x \rightarrow +\infty} P(x) = 100$$

Una nota
sull'istogramma

VOTO	Frequenza assoluta
60 I- 70	10
70 I- 80	21
80 I- 90	20
90 I- 100	28
	79

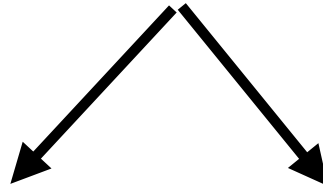


ETA'	Frequenza assoluta
30 I- 35	30
35 I- 38	80
38 I- 40	60
40 I- 44	30
	200



Dettagli sull'istogramma

tabelle di frequenza per classi di modalità



di uguale ampiezza

si possono usare direttamente le frequenze sull'asse delle ordinate
(naturalmente non è sbagliato usare le densità di frequenza)

$$d_i = \frac{n_i}{A_i}$$

$$d_i = \frac{f_i}{A_i}$$

$$d_i = \frac{p_i}{A_i}$$

di differente ampiezza

si devono usare le densità di frequenza sull'asse delle ordinate
(è sbagliato usare le densità di frequenza)

nel caso di classe di uguale ampiezza tutte le A_i sono uguali

Dettagli sull'istogramma

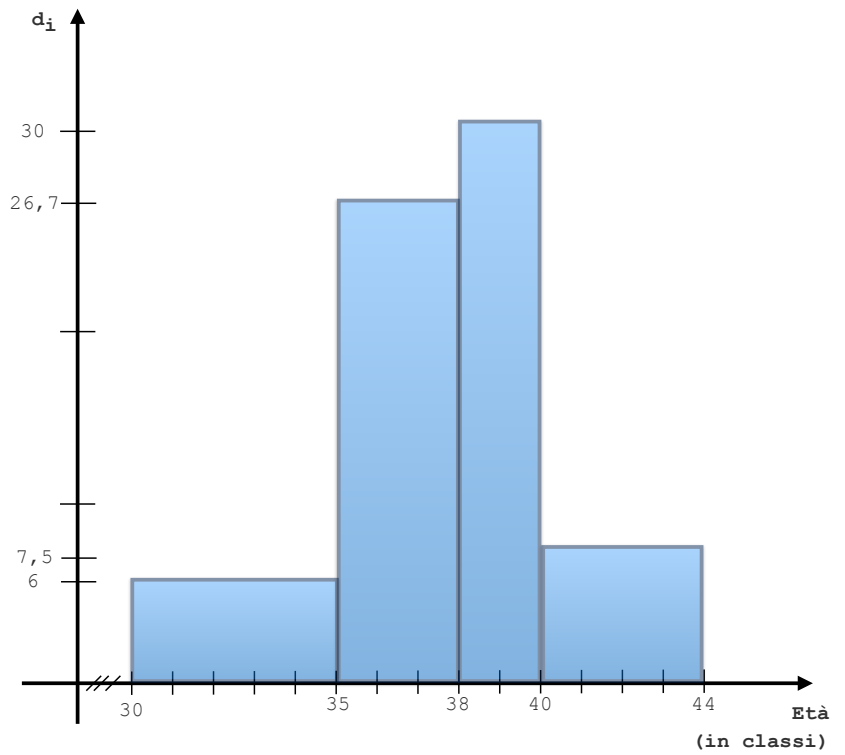
L'istogramma è una rappresentazione areale: l'area di ciascun rettangolo è uguale (nel caso si usino le densità) o proporzionale (nel caso si usino le frequenze) alla frequenza della classe corrispondente

BASE x ALTEZZA

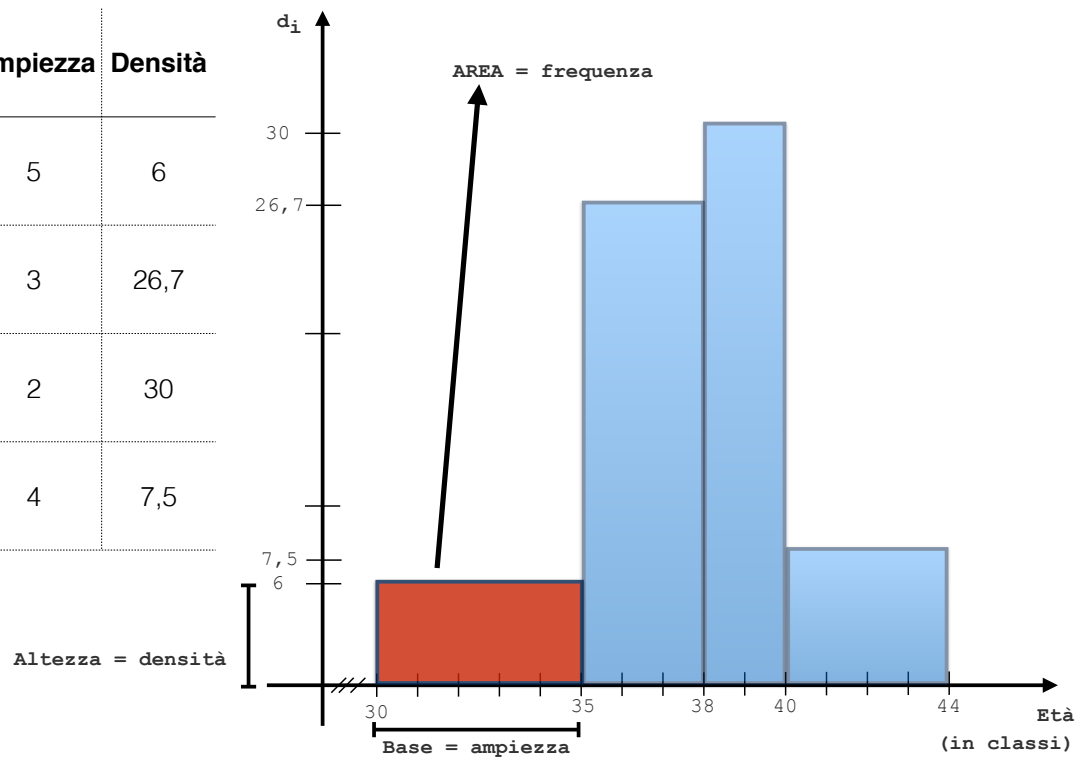
$$x_i - x_{i-1} \quad \times \quad d_i$$

$$A_i \quad \times \quad \frac{\text{frequenza}}{A_i} = \text{frequenza} \\ (n_i, f_i, p_i)$$

ETA'	Frequenza assoluta	Ampiezza	Densità
30 I- 35	30	5	6
35 I- 38	80	3	26,7
38 I- 40	60	2	30
40 I- 44	30	4	7,5
	200		



ETA'	Frequenza assoluta	Ampiezza	Densità
30 I- 35	30	5	6
35 I- 38	80	3	26,7
38 I- 40	60	2	30
40 I- 44	30	4	7,5
	200		

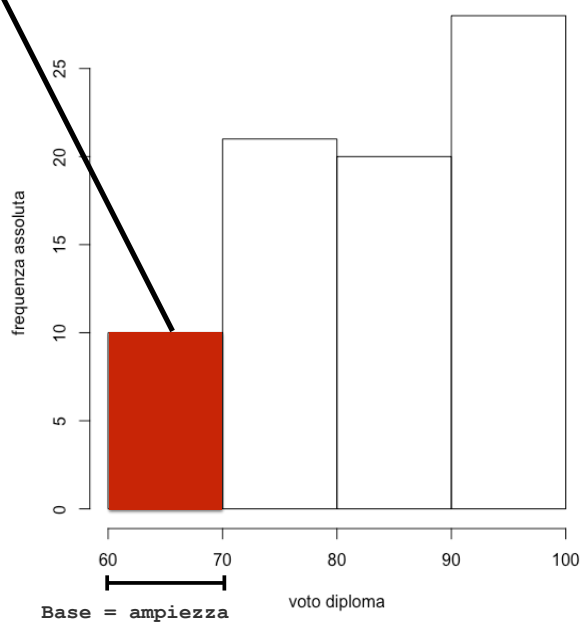


VOTO	Frequenza assoluta
60 I- 70	10
70 I- 80	21
80 I- 90	20
90 I- 100	28
	79

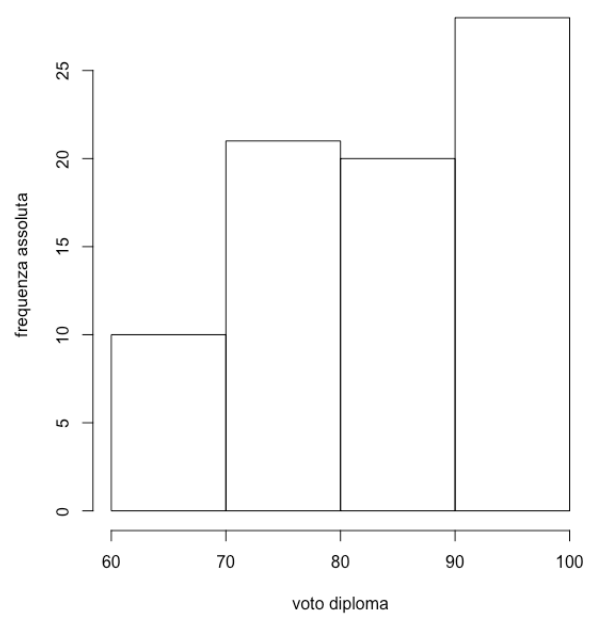
AREA proporzionale alla frequenza

Distribuzione voto di diploma

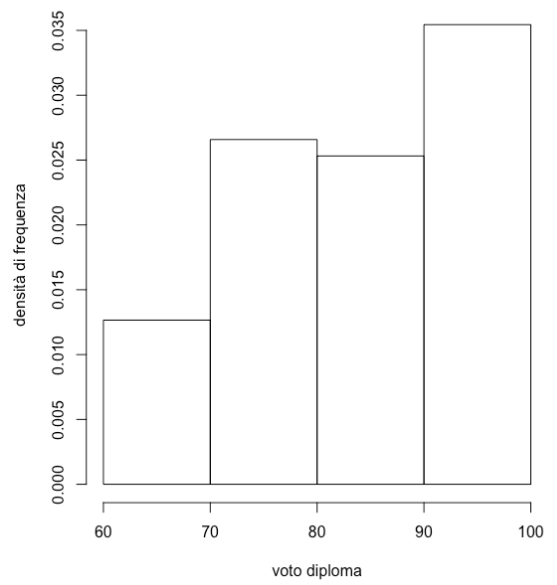
Altezza = frequenza



Distribuzione voto di diploma



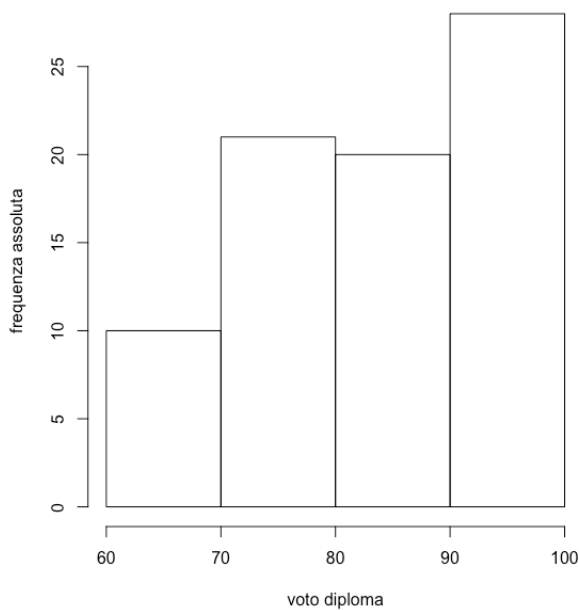
Distribuzione voto di diploma (istogramma normalizzato)



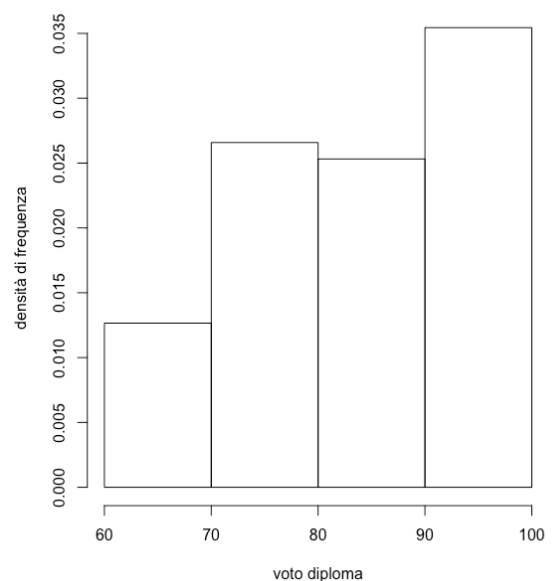
Quanto vale la somma delle aree di tutti i rettangoli?

Una nota sull'istogramma

Distribuzione voto di diploma



Distribuzione voto di diploma (istogramma normalizzato)



Quanto vale la somma delle aree di tutti i rettangoli?

Dettagli sull'istogramma

L'istogramma è una rappresentazione areale: l'area di ciascun rettangolo è uguale (nel caso si usino le densità) o proporzionale (nel caso si usino le frequenze) alla frequenza della classe corrispondente

BASE x ALTEZZA

$$x_i - x_{i-1} \times d_i$$

$$A_i \times \frac{\text{frequenza}}{A_i} = \text{frequenza} \\ (n_i, f_i, p_i)$$

frequenza $\rightarrow \sum_i \text{area}_i = n$

Classi di uguale ampiezza

densità $\rightarrow \sum_i \text{area}_i \propto n$

Classi di diversa ampiezza

PRINCIPALI RAPPRESENTAZIONI GRAFICHE
(vedi il file html con gli esempi + studio autonomo)

Variabili qualitative

Diagramma a barre, diagramma a torta, diagramma di Pareto

Variabili quantitative

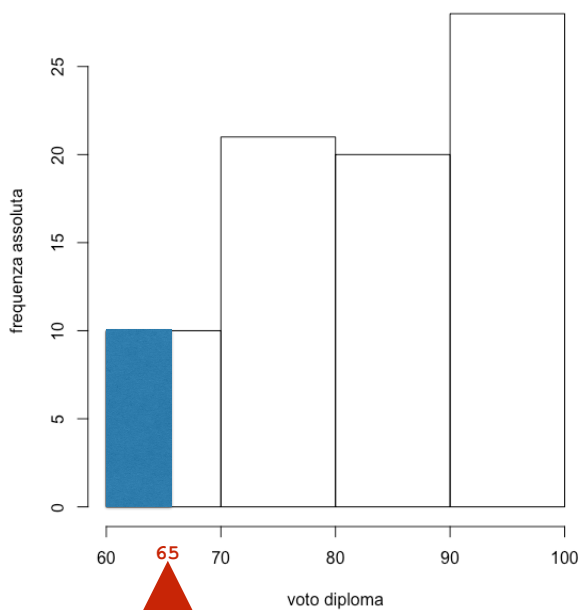
Diagramma ad aste (a bastoncini), dotplot, istogramma, poligono di frequenza, funzione di ripartizione empirica

STUDIO AUTONOMO

VOTO	Frequenza assoluta
60 I- 70	10
70 I- 80	21
80 I- 90	28
90 I- 100	28
	79

Nel caso di istogramma costruito con le frequenze possiamo ragionare con una semplice proporzione a partire dalla tabella

Distribuzione voto di diploma

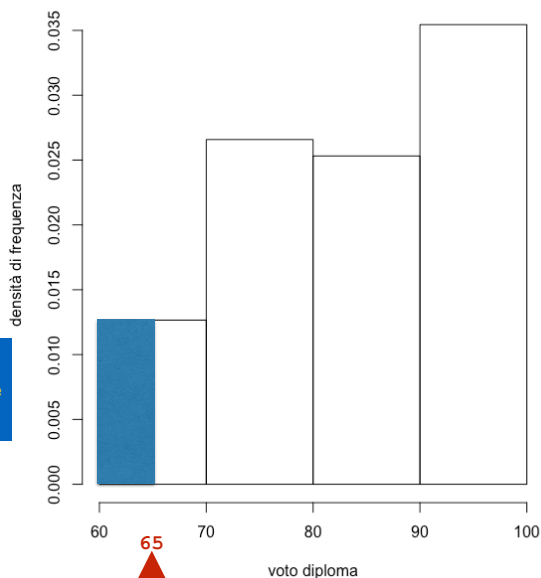


Quanti sono (approssimazione) gli studenti che hanno avuto un voto minore di 65?

VOTO	Frequenza assoluta
60 I- 70	10
70 I- 80	21
80 I- 90	28
90 I- 100	28
	79

Nel caso di istogramma costruito con le densità possiamo calcolare l'area del rettangolo

Distribuzione voto di diploma (istogramma normalizzato)

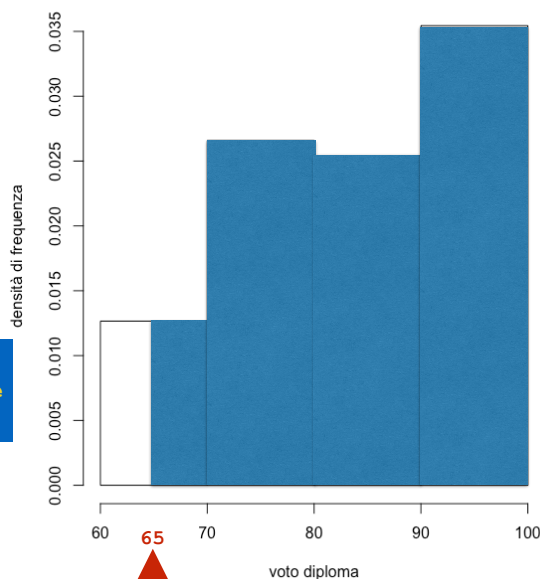


Quanti sono (approssimazione) gli studenti che hanno avuto un voto minore di 65?

VOTO	Frequenza assoluta
60 I- 70	10
70 I- 80	21
80 I- 90	28
90 I- 100	28
	79

Nel caso di istogramma costruito con le densità possiamo calcolare l'area del rettangolo

Distribuzione voto di diploma (istogramma normalizzato)



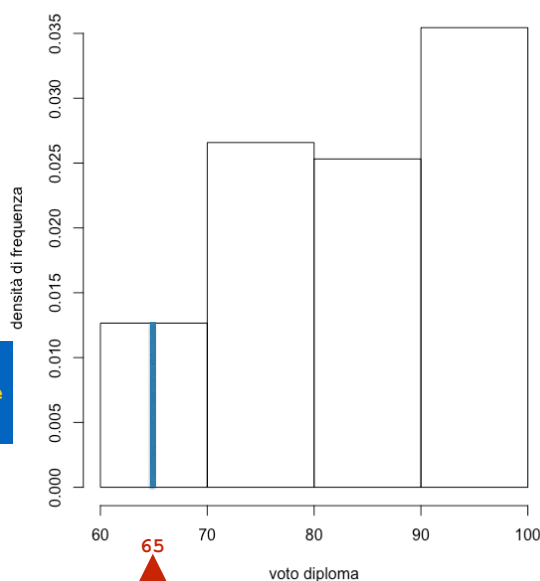
Quanti sono (approssimazione) gli studenti che hanno avuto un voto maggiore di 65?

“Domandona”
(pensateci bene)

VOTO	Frequenza assoluta
60 I- 70	10
70 I- 80	21
80 I- 90	28
90 I- 100	79

Nel caso di istogramma costruito con le densità possiamo calcolare l'area del rettangolo

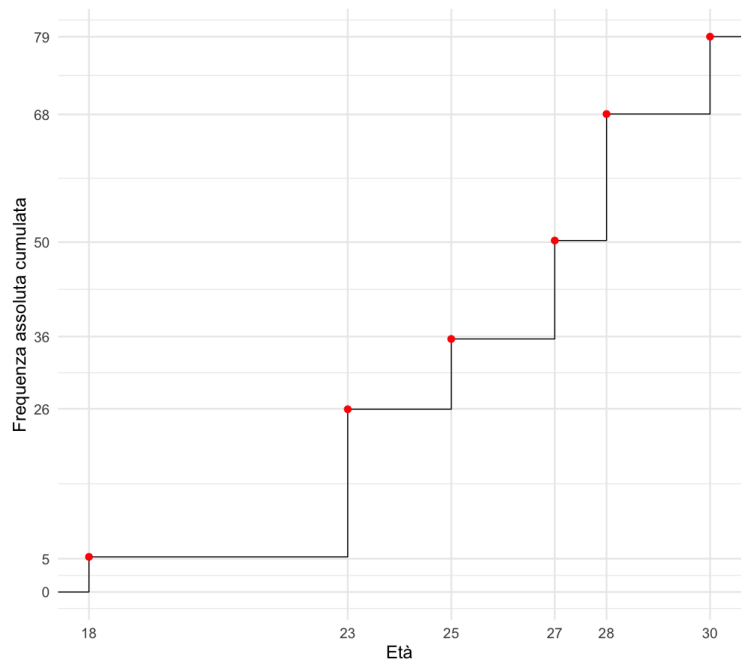
Distribuzione voto di diploma (istogramma normalizzato)



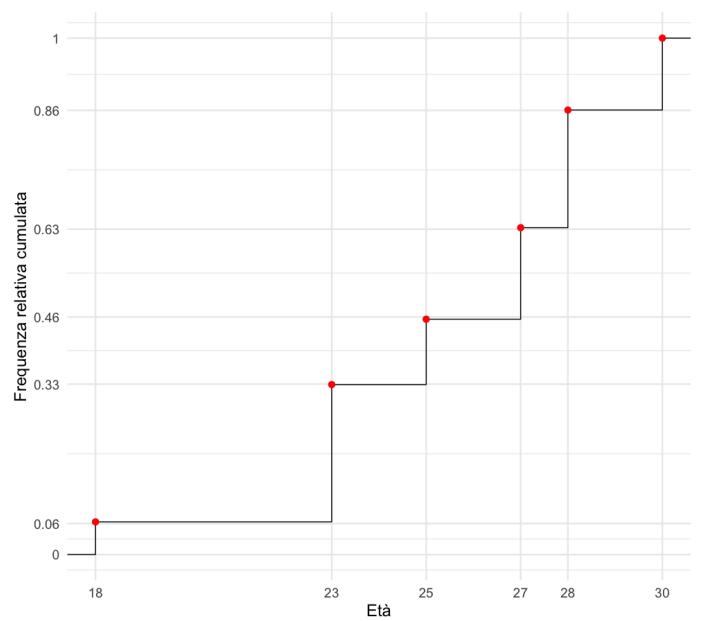
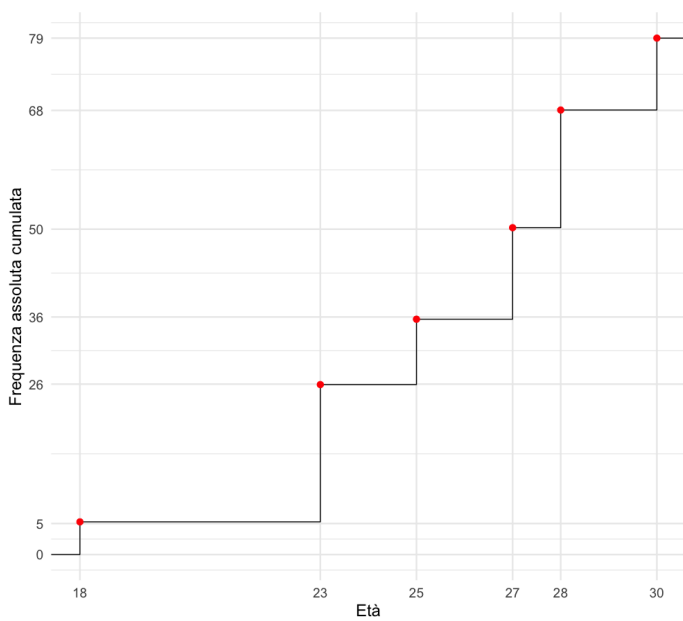
Quanti sono (approssimazione) gli studenti che hanno avuto un voto esattamente uguale a 65?

Una nota sulla
funzione di
ripartizione
empirica

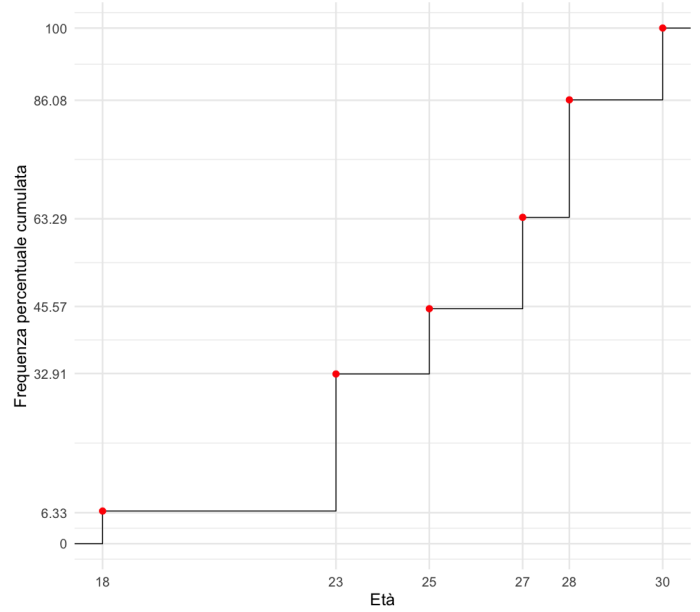
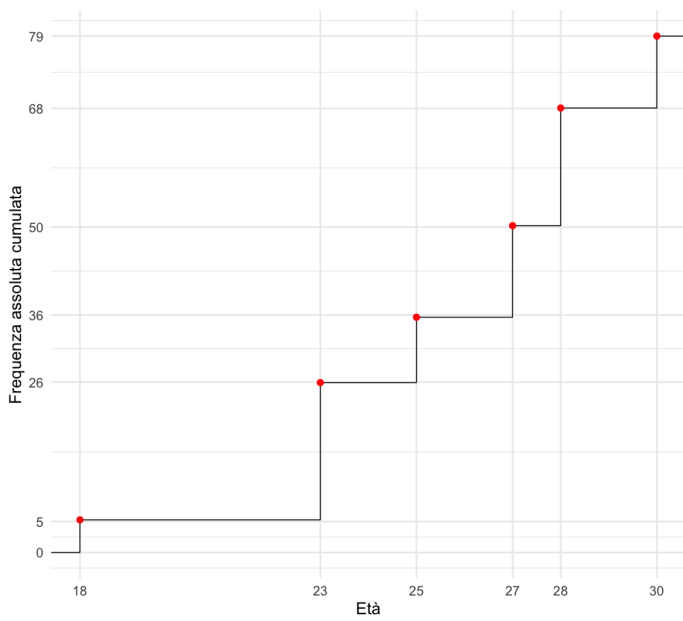
VOTO	Frequenza assoluta	Frequenza cumulata
18	5	5
23	21	26
25	10	36
27	14	50
28	18	68
30	11	79
	79	



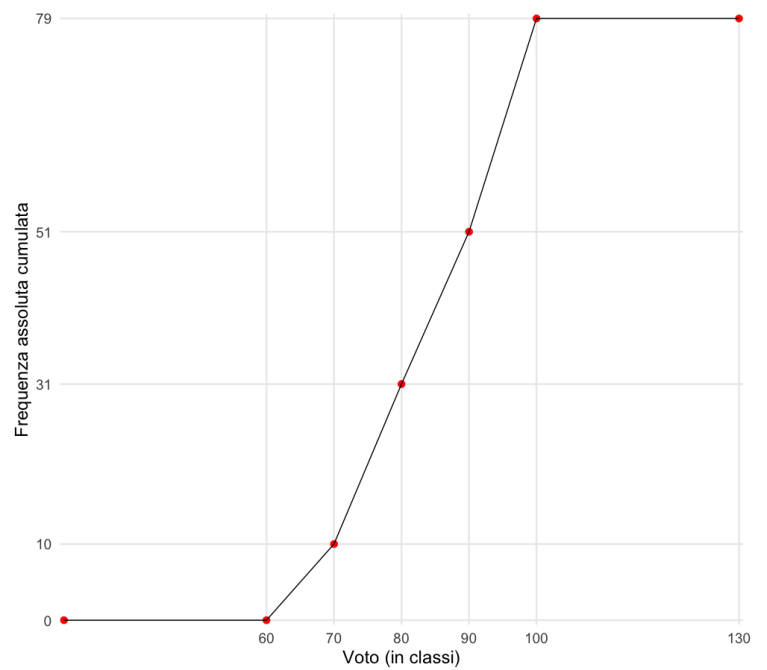
Trova le differenze...



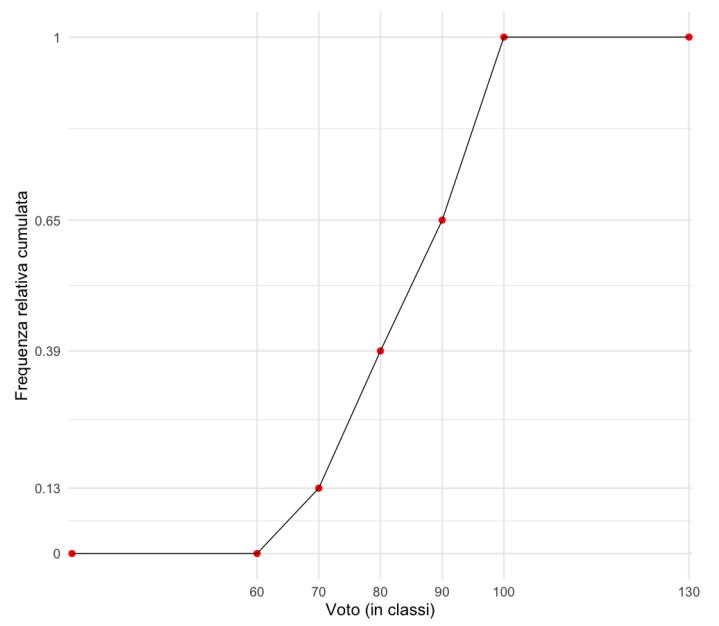
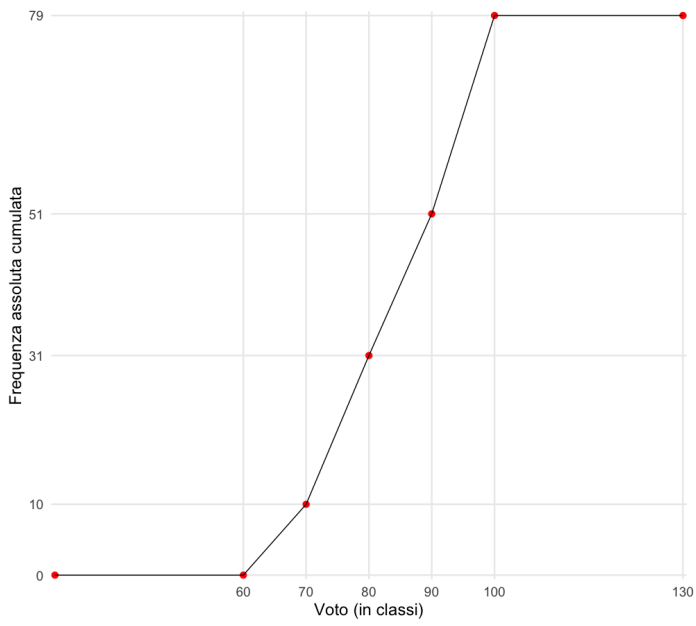
Trova le differenze...



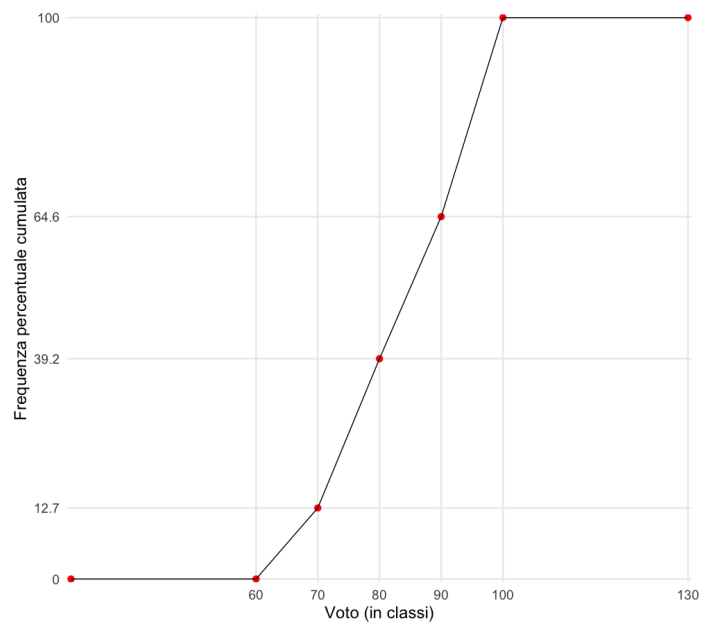
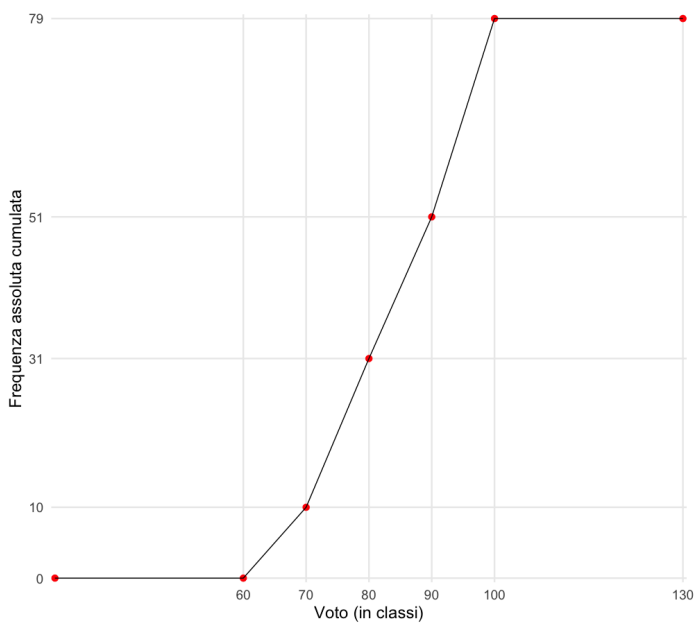
VOTO	Frequenza assoluta	Frequenza cumulata
60 I- 70	10	10
70 I- 80	21	31
80 I- 90	20	51
90 I- 100	28	79
	79	



Trova le differenze...

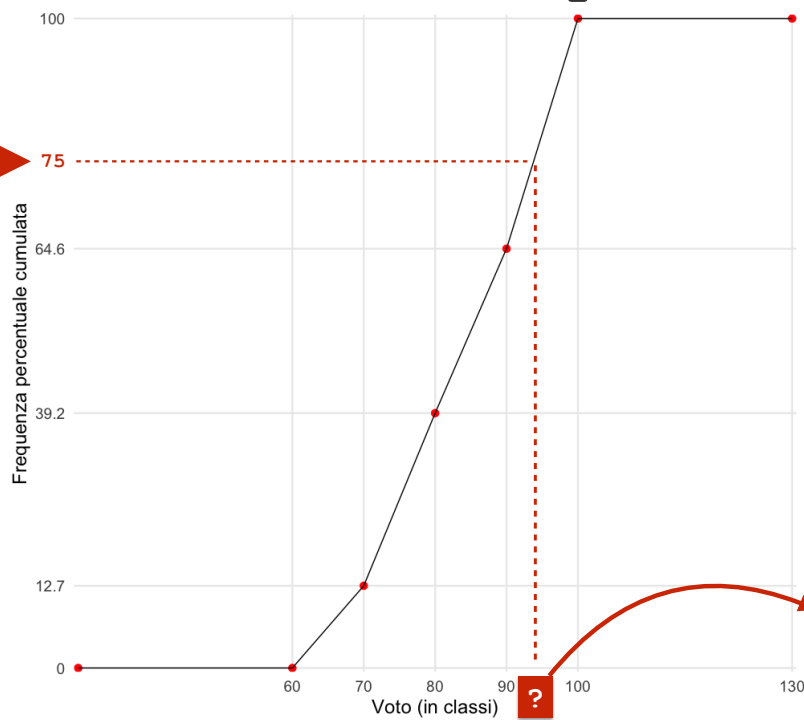


Trova le differenze...



Giochiamo un poco?

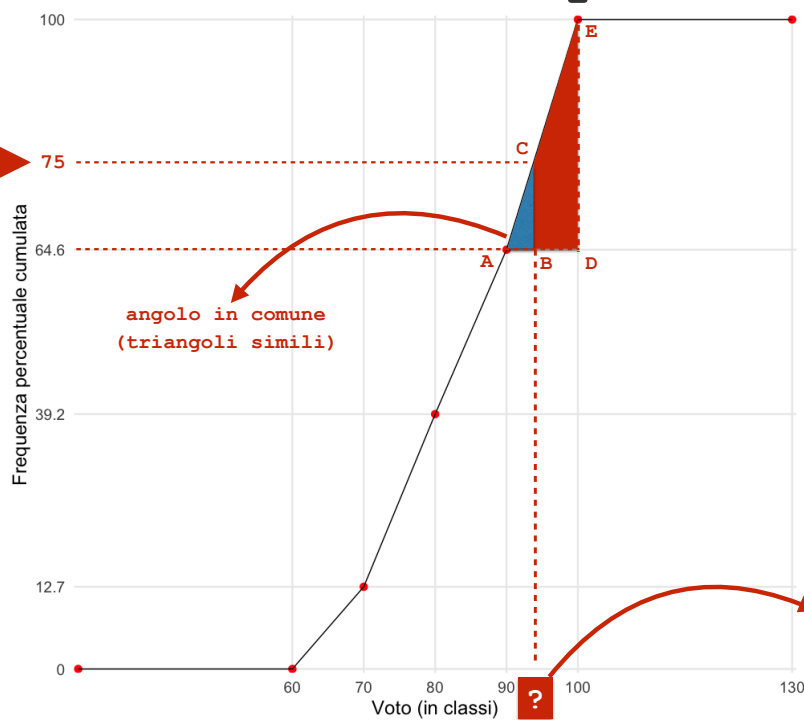
Qual è il voto superato solo dal 25% degli studenti?



75° percentile

Giochiamo un poco?

Qual è il voto superato solo dal 25% degli studenti?

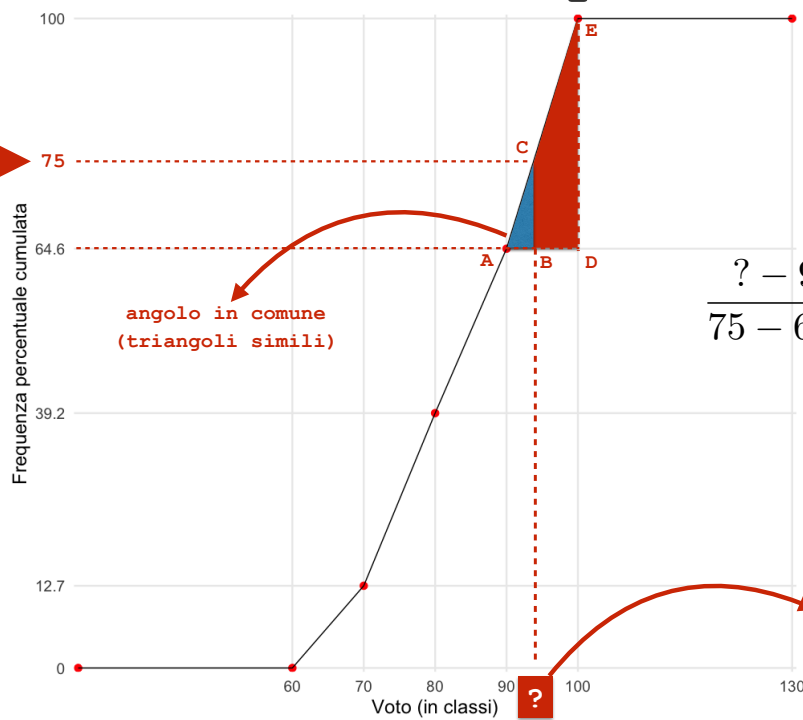


$$\frac{\overline{AB}}{\overline{BC}} = \frac{\overline{AD}}{\overline{DE}}$$

75° percentile

Giochiamo un poco?

Qual è il voto superato solo dal 25% degli studenti?



angolo in comune
(triangoli simili)

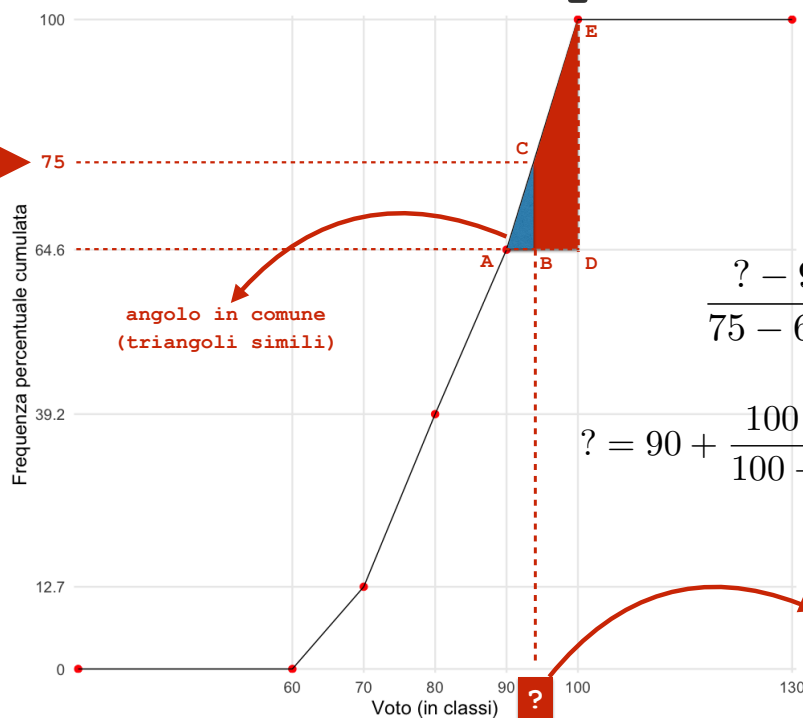
$$\frac{\overline{AB}}{\overline{BC}} = \frac{\overline{AD}}{\overline{DE}}$$

$$\frac{? - 90}{75 - 64.6} = \frac{100 - 90}{100 - 64.6}$$

75° percentile

Giochiamo un poco?

Qual è il voto superato solo dal 25% degli studenti?



angolo in comune
(triangoli simili)

$$\frac{\overline{AB}}{\overline{BC}} = \frac{\overline{AD}}{\overline{DE}}$$

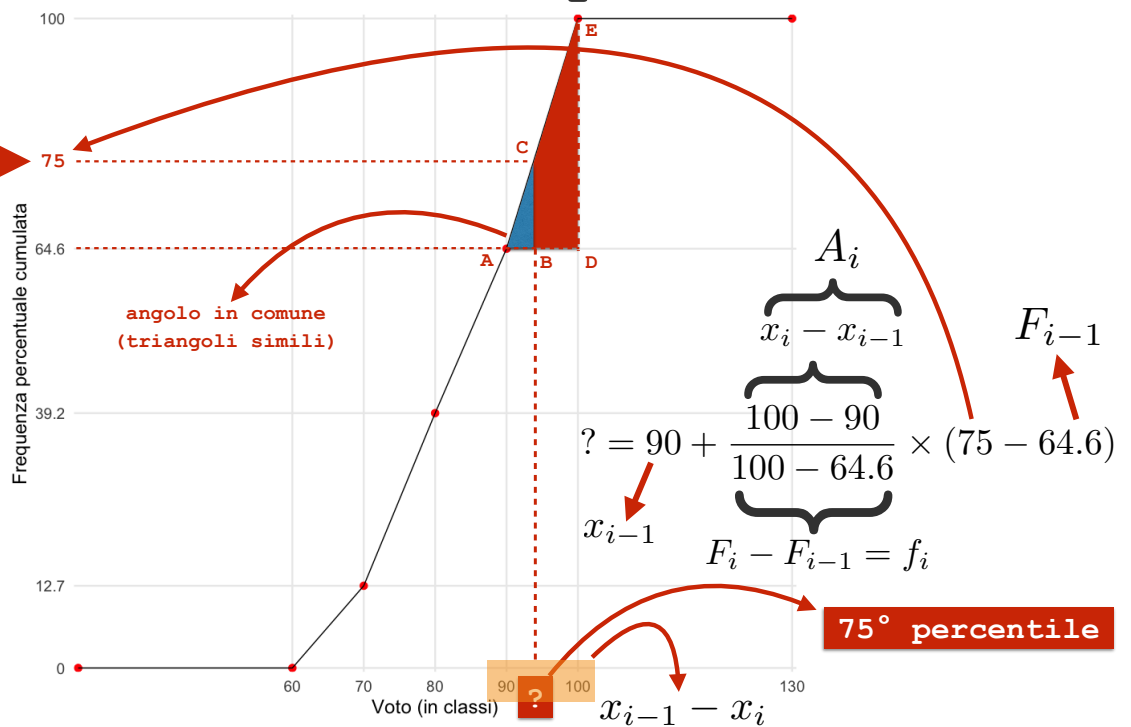
$$\frac{? - 90}{75 - 64.6} = \frac{100 - 90}{100 - 64.6}$$

$$? = 90 + \frac{100 - 90}{100 - 64.6} \times (75 - 64.6)$$

75° percentile

Giochiamo un poco?

Qual è il voto superato solo dal 25% degli studenti?



In sintesi...

$$q_{0.75} = x_{i-1} + \frac{x_i - x_{i-1}}{p_i} (75 - P_{i-1}) = q_3$$

$100 \times \frac{75}{100}$

$$q_{0.75} = x_{i-1} + \frac{x_i - x_{i-1}}{n_i} \left(n \frac{75}{100} - N_{i-1} \right) = q_3$$

$$q_{0.75} = x_{i-1} + \frac{x_i - x_{i-1}}{f_i} (0.75 - F_{i-1}) = q_3$$

$1 \times \frac{75}{100} = \frac{3}{4}$

In generale...

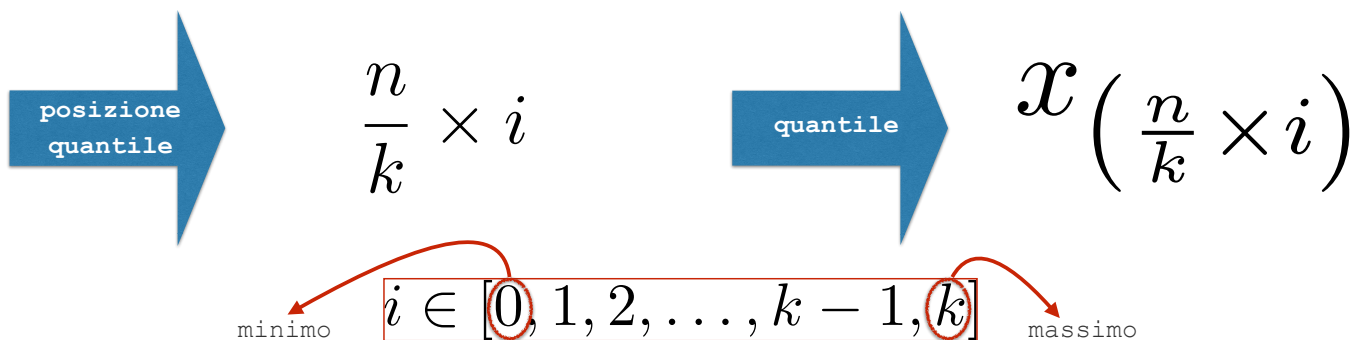
$$\tau \in [0, 1]$$

$$q_\tau = x_{i-1} + \frac{x_i - x_{i-1}}{f_i} (\tau - F_{i-1})$$

$$q_\tau = x_{i-1} + \frac{x_i - x_{i-1}}{p_i} (\tau \times 100 - P_{i-1})$$

$$q_\tau = x_{i-1} + \frac{x_i - x_{i-1}}{n_i} (\tau \times n - N_{i-1})$$

Alternativa?



k è il numero di parti in cui vogliamo dividere i dati

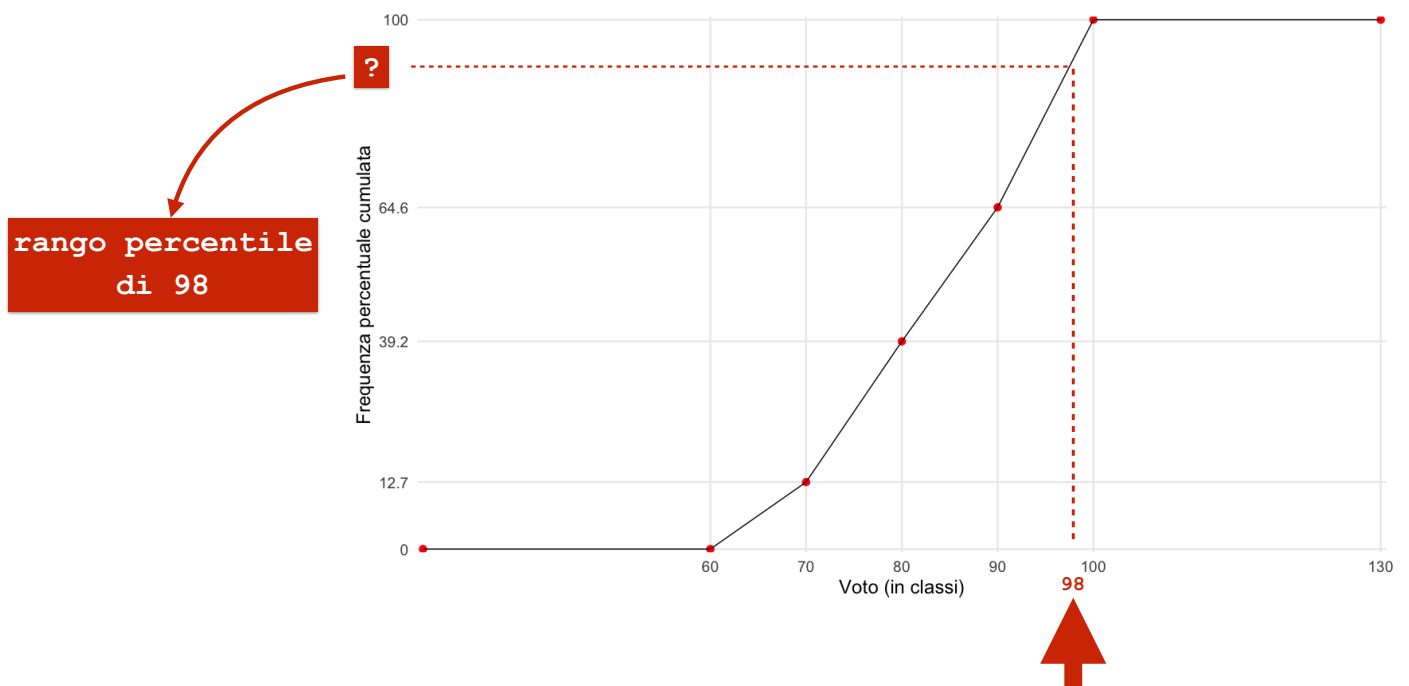
i è la parte che ci interessa

k = 4, i = 3 → Terzo quartile

k = 10, i = 9 → Nono decile

k = 100, i = 20 → Ventesimo percentile

“Domandona” (e adesso pensateci voi)



Giochiamo un poco?

Quanti sono gli studenti (in %) che hanno avuto un voto maggiore di 98?