POLO UNIVERSITARIO
FONDAZIONE
PORTOGRUARO CAMPUS

Corso di Statistica Sociale

CORSO DI LAUREA: SCIENZE DELL'EDUCAZIONE

DOCENTE: FRANCESCO SANTELLI

Cosa sono i dati in classe?

Si parte da dati di natura numerica: solitamente continua

Da molte modalità (*in teoria infinite*) a poche modalità (poche classi)

Questo processo prende anche il nome di discretizzazione

Dati in serie:

0, 12, 5, 6…

Classi	***Frequenza***
0-5	10
6-10	8
11-15	12
16-20	10

Da 21 modalità (tutti i numeri da 0 a 20) a 3 classi, quindi 3 modalità

Vantaggio: maggiore leggibilità e interpretabilità

Svantaggio: perdita parziale di informazione… Perché!?

Come costruire le classi

Esistono innumerevoli metodi. Gli approcci più utilizzati sono 3:

1) Costruzione di classi in base a una teoria che ne prestabilisce il significato. Ad esempio: fasce di reddito (pensate alle aliquote irpef), suddivisione dei voti agli esami (solitamente i voti da 0 a 17 sono un'unica classe), fasce di peso di bambini di una data età (sotto una certa soglia denutrizione, sopra una certa soglia rischio obesità ecc.).

Altri approcci sono basati direttamente sui dati (*data driven*):

2) Classi equiampie: sono classi che hanno tutte la stessa *ampiezza* (a meno di arrotondamenti e piccole approssimazioni). Esempio: dividiamo i voti all'esame da in 3 classi: 0-10, 11-20, 21-30.

Domanda: ampiezza di queste classi?

3) Classi equifrequenti: sono classi che hanno tutte la stessa *frequenza* (a meno di arrotondamenti e piccole approssimazioni). Esempio: dividiamo 30 studenti in 3 classi in modo tale che ogni classe abbia lo stesso numero di studenti al suo interno.

Domanda: frequenza di queste classi?

Esercizietto 1

0	11
0	11
1	11
1	12
2	13
2	14
2	15
3	15
4	15
5	15
6	16
6	16
7	16
8	17
9	18
9	19
10	20
10	20
11	20
11	20

Da dati in serie ad istogramma!! (sono già ordinati, altrimenti si devono ordinare! Date le seguenti classi:

Completare la seguente tabella di frequenza associata:

Classi	***Frequenza***	***Ampiezza***	***Densità di Frequenza***
0-5			
6-10			
11-15			
16-20			

Esercizietto 2

0	11
0	11
1	11
1	12
2	13
2	14
2	15
3	15
4	15
5	15
6	16
6	16
7	16
8	17
9	18
9	19
10	20
10	20
11	20
11	20

1) Costruire i dati in classe utilizzando 4 classi equiampie

Classi	***Frequenza***	***Ampiezza***	***Densità di Frequenza***
?			
?			
?			
?			

2) Costruire i dati in classe utilizzando 4 classi equifrequenti

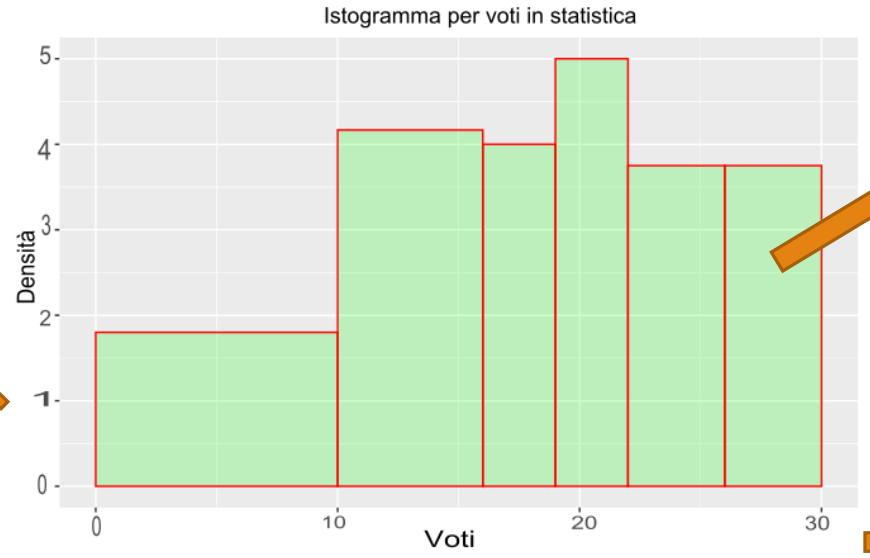
Classi	***Frequenza***	***Ampiezza***	***Densità di Frequenza***
?			
?			
?			
?			

Come rappresentare dati in classe?

Come detto, i dati in classe erano originariamente numerici, ora sono diventati in qualche modo discreti

L'istogramma è il modo migliore per rappresentare questa condizione. Esso tiene conto della natura originaria dei dati ma anche della loro nuova rappresentazione discreta.

Asse y, nuova quantità:
Densità di frequenza



Area di ogni rettangolino:
La frequenza, cioè quanti
Studenti ci sono dentro

Asse X: i voti, le classi

Diversi istogrammi possibili su stessi dati

Ciò che cambia è il numero di rettangolini. Essi sono associati al numero di classi identificate

Maggior numero di classi, sempre più rettangolini, che diventano sempre più «piccoli»

Poche classi, pochi rettangolini ma molto «grossi»

Quale rappresentazione è migliore? Molti rettangolini (classi) o pochi?

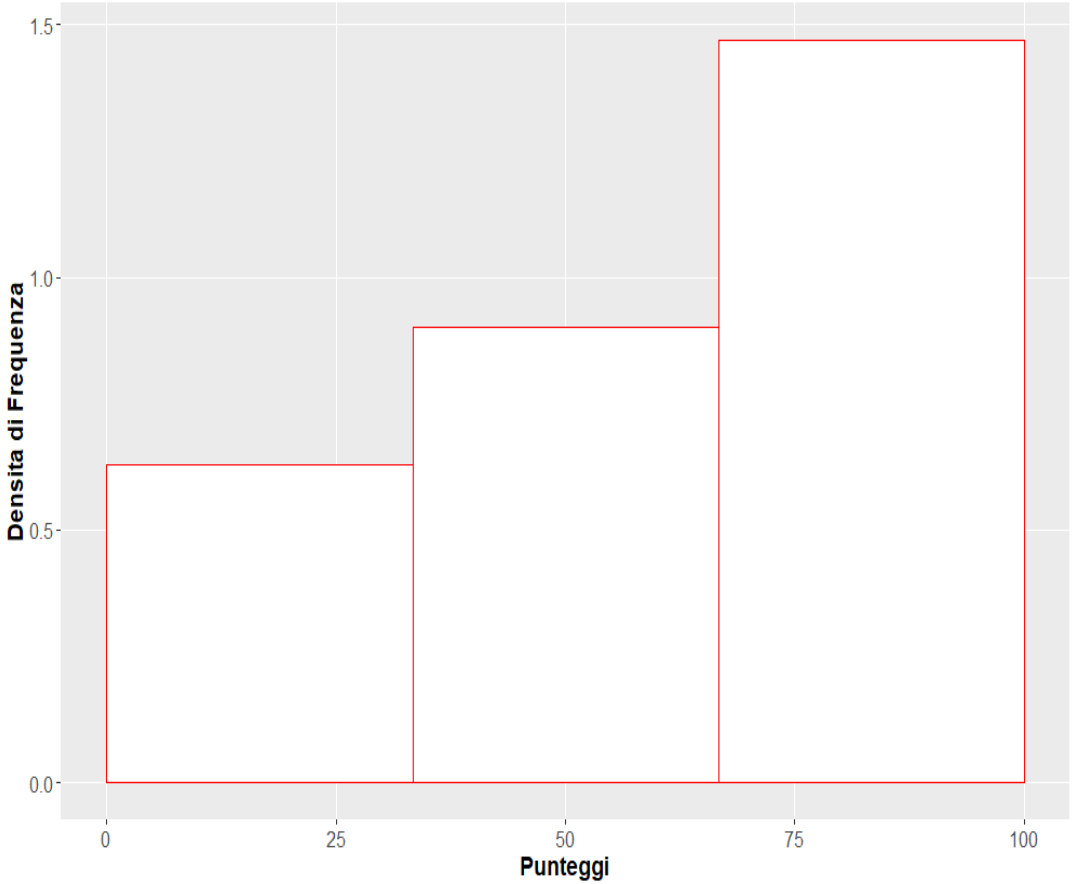
Non esiste una regola precisa e univoca!

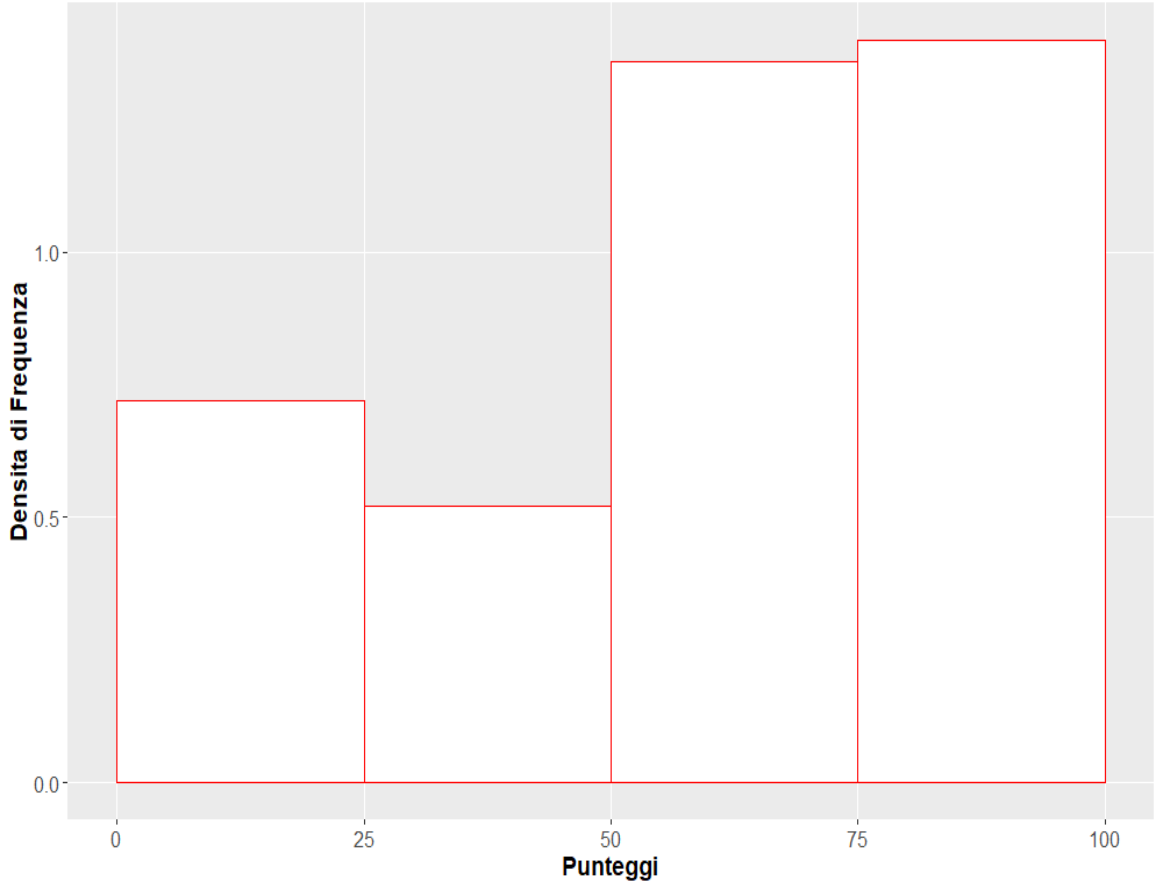
Dipende dal numero di osservazioni, da come i dati si distribuiscono ecc.

Comunque, l'istogramma deve essere informativo, cioè restituire una sorta di «verità» sulla distribuzione originaria dei dati numerici.

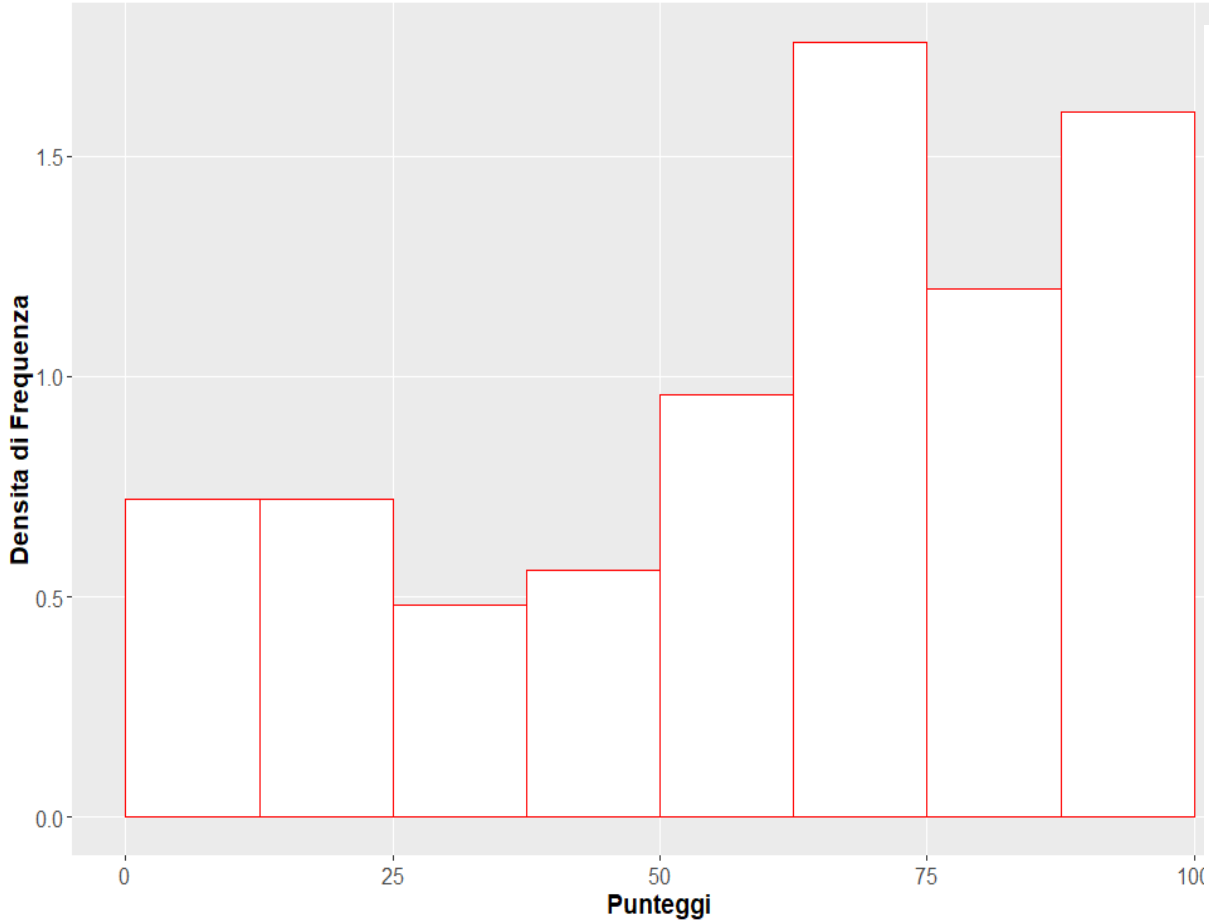
Numeri di rettangolini (classi) più utilizzati: 4, 5, 8, 10, 20.

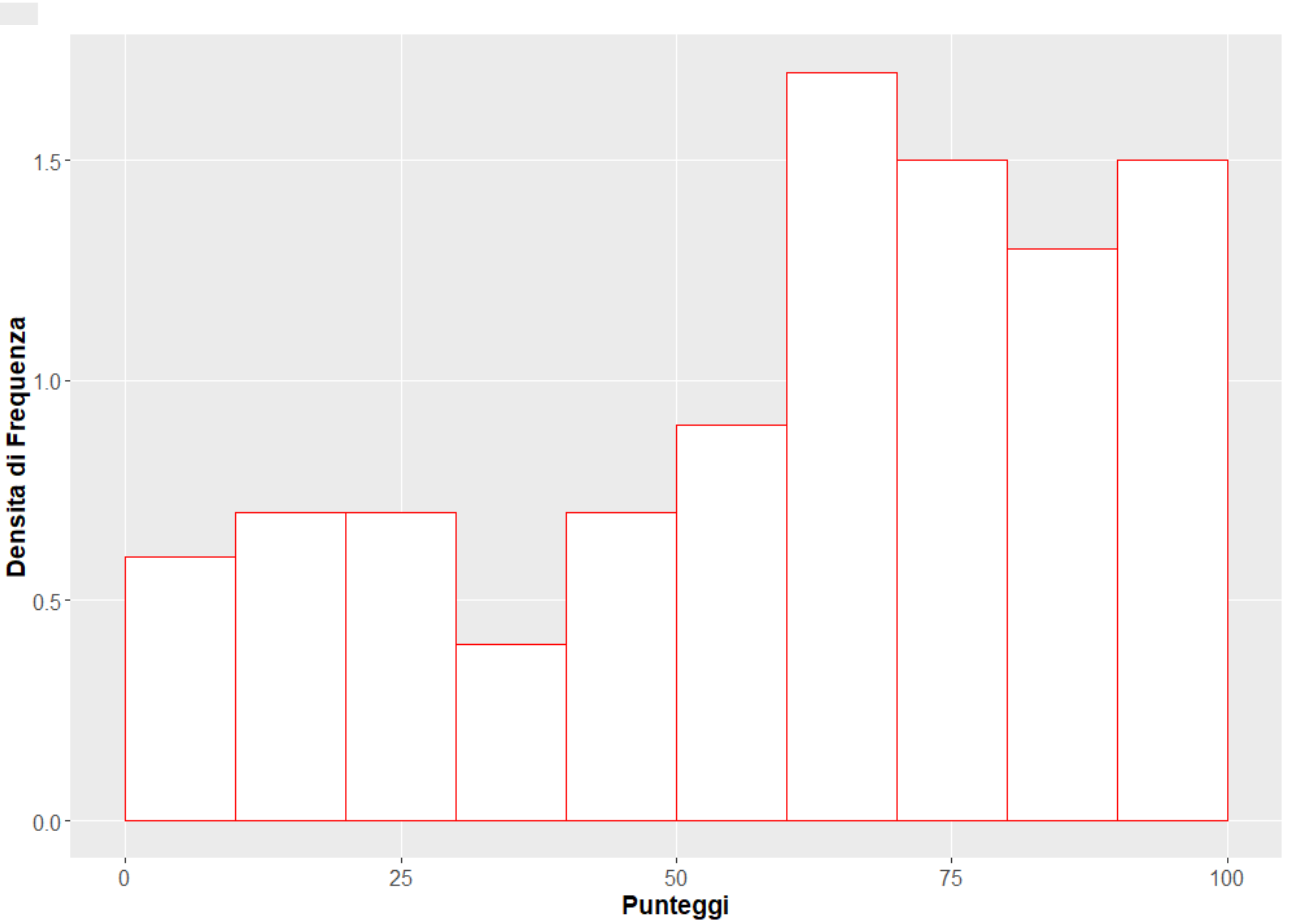
Esempi di istogrammi su stessi dati (1)



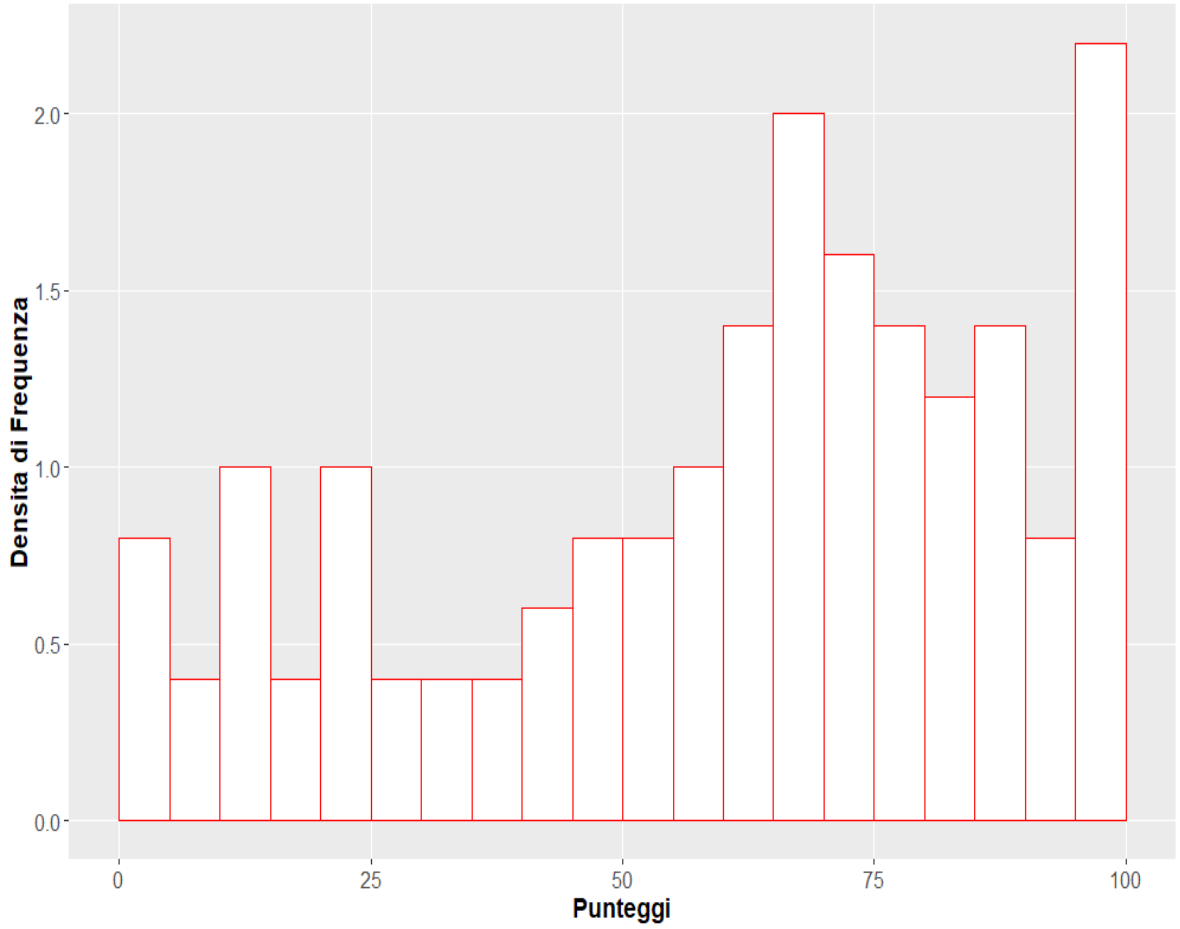


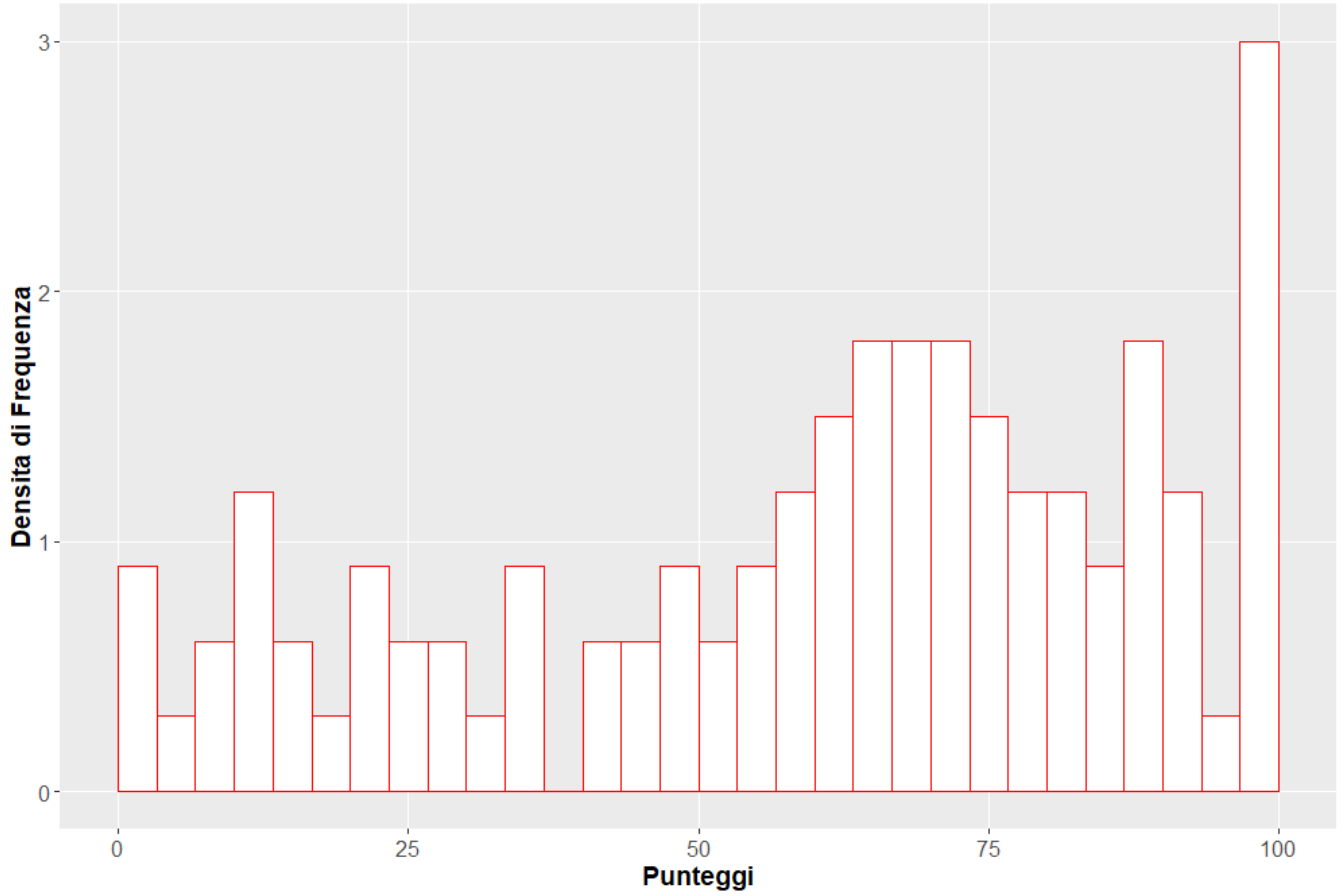
Esempi di istogrammi su stessi dati (2)





Esempi di istogrammi su stessi dati (3)





Media dati in classe

Quando passiamo da dati in serie a dati in classe perdiamo parte dell'informazione, ma possiamo comunque calcolare degli indici!

Regola generale: quando usiamo dati in classe, ogni classe viene rappresentata dal suo valore centrale, che chiameremo V.C.

1) ***Media dati in classe*** simile a media ponderata. Solo che a) al posto dei pesi ci sono le frequenze b) al posto delle x_i ci sono i valori centrali di ogni classe

Classi	***Frequenza***
0-5	10
6-10	8
11-15	12
16-20	10

$$\overline{x} = \frac{\sum_{i=i}^{C} V.C._i * F_i}{N}$$

Classi	***Frequenza***	***V.C.***	***V.C.*F***
0-5	10	2,5	25
6-10	8	8	64
11-15	12	13	156
16-20	10	18	180
			425
			10,625

Moda dati in classe

La moda è, come sempre, la modalità più frequente

In questo caso non utilizziamo le frequenze relative o percentuali o assolute, ma…

Utilizziamo la densità di frequenza, che ricordiamo essere pari a frequenza assoluta / ampiezza

Quindi ***Definizione*** la classe modale è la classe a cui è associata la densità di frequenza più alta

Classi	***Frequenza***	***V.C.***	***V.C.*F***	***Ampiezza***	***Densità di Frequenza***
0-5	10	2,5	25	5	2
6-10	8	8	64	4	2
11-15	12	13	156	4	3
16-20	10	18	180	4	2,5

Mediana per dati in classe è un calcolo troppo complesso che vi evito, ma se siete così bravi potete cimentarvi nel lavoro di gruppo (chiedete a ricevimento e vi sarà dato)