

Distribuzione di frequenza

Una **distribuzione di frequenza** è una tabella dove in corrispondenza delle modalità viene riportato il numero di volte che quelle stesse modalità si sono verificate.

Ad ogni modalità della variabile statistica viene associata la sua frequenza.

Modalità (voto)	Frequenze assolute (n. 26 studenti)
24	3
25	4
26	4
27	6
28	2
29	5
30	2

Distribuzione di frequenza

La distribuzione di frequenza può essere utilizzata anche per variabili non numeriche (es. regioni)

Regioni	Stranieri residenti
Piemonte	203,9
Valle d'Aosta-Vallée d'Aoste	3,7
Lombardia	577,3
Bolzano-Bozen	24,0
Trento	22,1

Si può raggruppare la **distribuzione di frequenza in classi**.

L'ampiezza delle classi è arbitraria, purchè si utilizzi un criterio di scelta che tenga conto dell'obiettivo dell'analisi.

Modalità (classi di voto)	Frequenze assolute (n. 26 studenti)
24-26	11
27-29	13
30	2

Frequenze relative

Per confrontare due distribuzioni è sufficiente guardare le frequenze assolute?

N. di componenti in famiglia	Campania Freq. assolute	Valle d'Aosta Freq. assolute
1	59	11
2	74	13
3	75	11
4	66	14
5	23	8
6	10	6
Totale	307	63

Frequenze relative

Per confrontare due distribuzioni è sufficiente guardare le frequenze assolute?

N. di componenti in famiglia	Campania		Valle d'Aosta	
	Freq. assolute	Freq. relative	Freq. assolute	Freq. relative
1	59	0,192	11	0,175
2	74	0,241	13	0,206
3	75	0,244	11	0,175
4	66	0,215	14	0,222
5	23	0,075	8	0,127
6	10	0,033	6	0,095
Totale	307	1	63	1

Frequenze relative

N. di componenti in famiglia	Campania			
	Freq. assolute	Freq. relative	Freq. percentuali	Freq. cumulate
1	59	0,192	19,2	19,2
2	74	0,241	24,1	43,3
3	75	0,244	24,4	67,7
4	66	0,215	21,5	89,2
5	23	0,075	7,5	96,7
6	10	0,033	3,3	100
Totale	307	1	100	

Tipo di frequenze

La **frequenza assoluta** rappresenta il numero intero di unità statistiche nelle quali è stata osservata una certa caratteristica.

Le **frequenze relative** si calcolano dividendo la frequenza assoluta di una determinata modalità o classe per il totale delle osservazioni. Si otterrà sempre un numero da 0 a 1.

Le frequenze relative possono essere moltiplicate per 100 ottenendo le **frequenze relative percentuali** (il totale è sempre 100%).

La **frequenza cumulata** di una modalità o classe è pari alla sua frequenza assoluta sommata alle frequenze assolute delle modalità che la precedono.

Le distribuzioni per le variabili continue

Nel caso di **variabili continue** dobbiamo tener conto del fatto che tra due modalità ce ne possono essere infinite altre. Il risultato probabile è che ci siano solo modalità distinte nel collettivo.

Ad esempio in questo caso:

	peso alla nascita
Michela	1.95
Francesco	2.15
Valentina	2.00
Giacomo	2.35
Matteo	3.10
Federica	2.45

Le distribuzioni per le variabili continue

Bisogna considerare le classi di modalità e la frequenza degli elementi nella classe così per la classe (x_{i-1}, x_i) si conteggiano tutte le modalità di X comprese in $x_{i-1} < X \leq x_i$.

La frequenza n_i mi dice quante unità presentano modalità comprese nell'intervallo (x_{i-1}, x_i) .

Come si calcola l'ampiezza delle classi?

Un criterio può essere prendere delle classi equi-ampie con ciascuna ampiezza uguale.

Tipo di frequenze

Supponiamo di voler sapere quanti dipendenti di una società abbiamo un reddito uguale o inferiore a 30.000€ l'anno.
Totale dei dipendenti = 120



Classi di reddito €	Frequenze assolute	Frequenze relative	Frequenze%	Frequenze relative cumulate	Frequenze % cumulate
< 15.000	16	0,13	13,3	0,13	13,3
15.000-30.000	64	0,53	53,3	0,67	66,7
30.000-45.000	30	0,25	25,0	0,92	91,7
45.000-60.000	8	0,07	6,7	0,98	98,3
60.000 <	2	0,02	1,7	1	100
	120	1	100		

Dalla tabella si evince che il 67% dei dipendenti ha un reddito uguale o inferiore a 30.000€.

Tipo di frequenze

Supponiamo ora di voler sapere quanti dipendenti della stessa società abbiamo un reddito superiore a 45.000€ l'anno.

Bisogna calcolare la **frequenza retrocumulata** sommando le frequenze a partire dall'ultima classe fino alla prima.

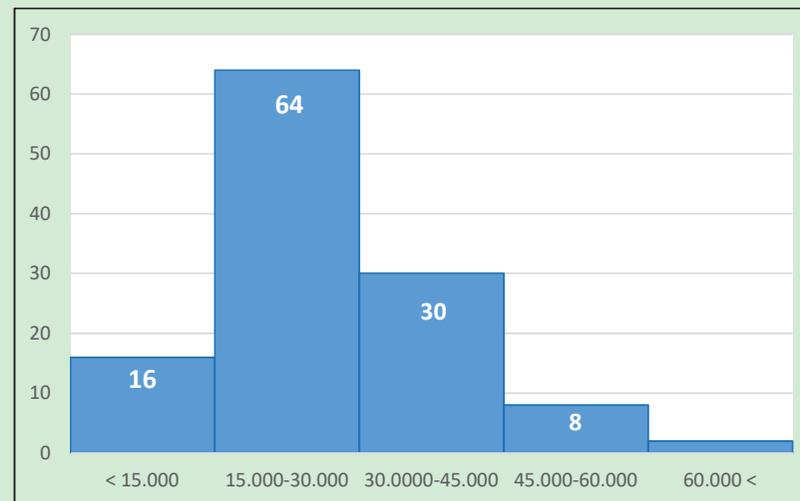
Classi di reddito €	Frequenze assolute	Frequenze assolute retrocumulate	Frequenze relative retrocumulate	Frequenze % retrocumulate
< 15.000	16	120	1	100
15.000-30.000	64	104	0,87	86,67
30.000-45.000	30	40	0,33	33,33
45.000-60.000	8	10	0,08	8,33
60.000 <	2	2	0,02	1,67
	120			

Dalla tabella si evince che 10 dipendenti hanno un reddito superiore a 45.000€, ossia l'8,33%.

Gli istogrammi

Gli **istogrammi** sono grafici che consistono in una serie di rettangoli con la base che poggia sull'asse orizzontale del piano cartesiano.

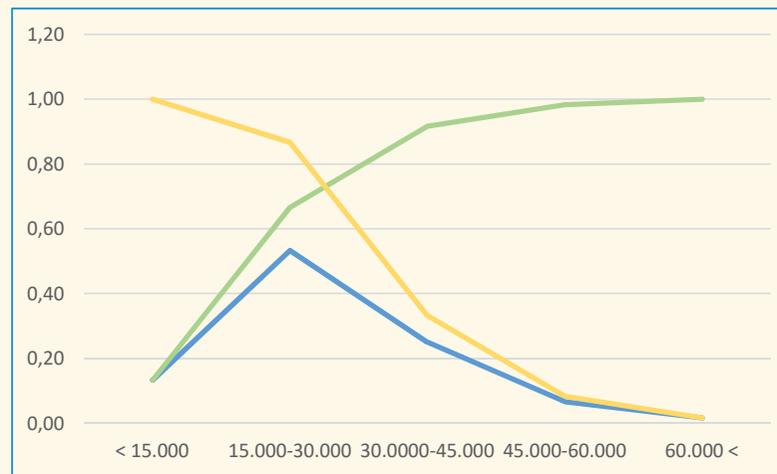
La base è data dall'ampiezza della classe, mentre l'altezza dipende dalla frequenza.



Le ogive

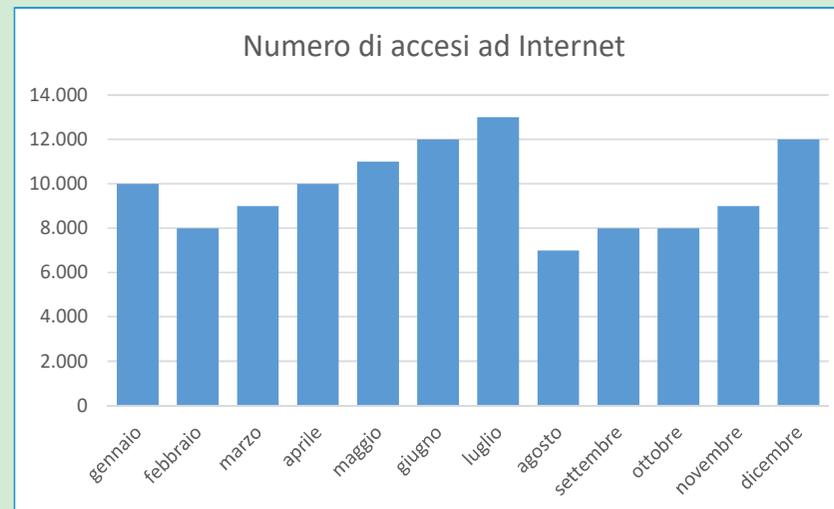
Le **ogive** sono utili per rappresentare le frequenze cumulate e retrocumulate.

Le ogive dal basso sono relative alle frequenze cumulate, le ogive dall'alto alle frequenze retrocumulate.



I diagrammi a barre

I **diagrammi a barre** sono simili agli istogrammi, ma vengono utilizzati per rappresentare dati numerici discreti o categorici. Sono utilizzati per visualizzare i raffronti tra categorie. Non c'è continuità tra le classi per cui le barre sono distanziate le une dalle altre.



Le serie storiche

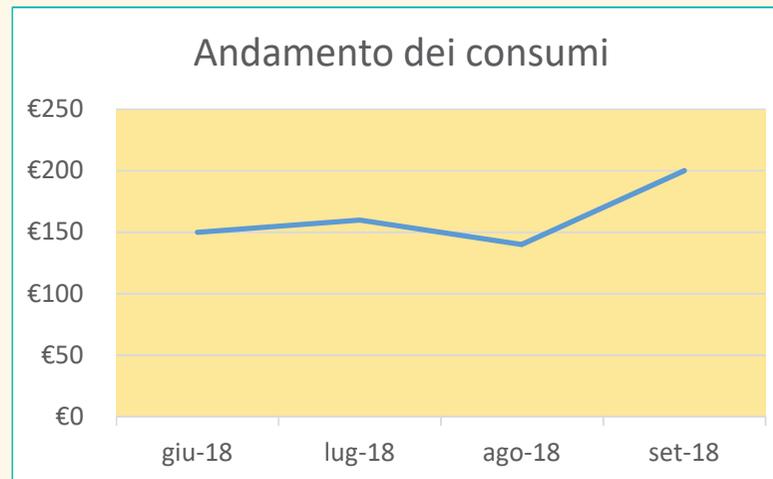
La **serie storica** rappresenta la dinamica di un certo fenomeno nel tempo.

Supponiamo di avere una variabile, un'unità ed il tempo: i consumi in Italia.

Per l'Italia (unità) abbiamo raccolto le informazioni per ciascun mese (tempo) dei consumi (variabile).

In questo caso abbiamo la cosiddetta **serie storica dei consumi**.

Grafico a linee



Calcolo della media per una distribuzione di frequenza

Modalità di x	Frequenze assolute (ni)
38	1
39	1
40	3
41	4
42	8
43	1
44	0
45	1
46	1
	20

$$\mu = \frac{\sum_{i=1}^N x_i n_i}{\sum_{i=1}^N n_i}$$

$$\mu = \frac{(38*1+39*1+...+46*1)}{1+1+3...+1}$$

$$\mu = 41,55$$

Calcolo della media per una distribuzione di frequenza

Modalità di x	Frequenze relative	$x_i \cdot f_i$
38	0,05	1,9
39	0,05	2,0
40	0,15	6,0
41	0,20	8,2
42	0,40	16,8
43	0,05	2,2
44	0,00	0,0
45	0,05	2,3
46	0,05	2,3
	1	41,55

Esercizio

Calcolare le frequenze relative e il numero medio di macchine possedute per famiglia

Macchine possedute (x_i)	N. di famiglie	
0	35	
1	50	
2	66	
3	15	

Esercizio

Calcolare le frequenze relative e il numero medio di macchine possedute per famiglia

Macchine possedute (x_i)	N. di famiglie	(f_i)	$x_i * f_i$
0	35	0,21	0,00
1	50	0,30	0,30
2	66	0,40	0,80
3	15	0,09	0,27
		1	1,36

Calcolo della media per una distribuzione di frequenza per classi

Nel caso in cui la distribuzione sia per classi occorre fare un passaggio ulteriore, calcolare il valore centrale di ciascuna classe:

$$\frac{(x_i + x_{i-1})}{2}$$

Ipotesi di **equidistribuzione**, ossia che i valori siano distribuiti uniformemente all'interno della classe

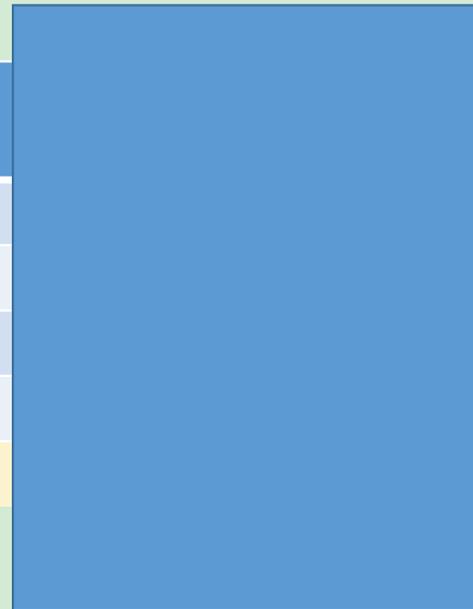
Classi di modalità	Frequenze assolute (ni)	Valore centrale della classe	ni*xi
0-60	5	30	150
60-120	27	90	2.430
120-180	255	150	38.250
180-300	571	240	137.040
300-600	312	450	140.400
600-1022	22	811	17.842
	1.192		336.112

$$\mu = 336.112 / 1.192 = 282$$

Esercizio

Calcolare la media per la seguente distribuzione per classi

x_i	n_i	
20-30	17	
30-40	65	
40-50	20	
50-60	35	
	137	



Esercizio

Calcolare la media per la seguente distribuzione per classi

xi	ni	Valori centrali	xi*ni
20-30	17	25	425
30-40	65	35	2.275
40-50	20	45	900
50-60	35	55	1.925
	137		5.525

$$\mu = 5.525/137 = 40,32$$

Calcolo della mediana

1- Disporre i dati in ordine crescente

2- Calcolare le frequenze cumulate

2- Calcolare l'indice i

oppure

2- Calcolare

$$i = \left(\frac{p}{100} \right) * n$$

$p = p$ -mo percentile

Se i non è intero, bisogna arrotondare per eccesso. L'intero più grande definisce la posizione del p -mo percentile. Se i è intero, il p -mo è la media delle posizioni i e $i+1$.

$$\text{Me} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{per } n \text{ dispari} \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} & \text{per } n \text{ pari} \end{cases}$$

Calcolo della mediana per una distribuzione di frequenza

Occorre calcolare le **frequenze assolute cumulate**.

$$\mu_e: \left(\frac{50}{100}\right) * 777 = 388,5 \rightarrow 389$$

$$\text{Oppure: } \frac{777+1}{2} = 389$$

$$\text{Oppure: } \frac{777}{2} = 388,5 \rightarrow 389$$

$\mu_e \rightarrow$

Modalità di x	Frequenze assolute	Frequenze cumulate
38	4	4
39	12	16
40	53	69
41	107	176
42	212	388
43	178	566
44	123	689
45	68	757
46	18	775
47	1	776
49	1	777
	777	

Calcolo della mediana per una distribuzione di frequenza per classi

Occorre calcolare le **frequenze assolute cumulate**.

$$Me = \frac{50}{100} * 75 = 37,5 \rightarrow 38$$

Oppure essendo un numero dispari $\frac{75+1}{2}=38$

La classe che contiene il valore mediano è la 300-400. **Me** è il valore centrale 350.

Classi di modalità	Frequenze assolute (ni)	Frequenze assolute cumulate
100-200	15	15
200-300	20	35
300-400	27	62
400-500	13	75
	75	



Calcolo della mediana per una distribuzione di frequenza per classi

Occorre calcolare le **frequenze assolute cumulate**.

$$Me = \frac{50}{100} * 76 = 38 \text{ e } 39 \rightarrow 38,5$$

$$\text{Oppure essendo un numero pari} = \frac{(76/2) + ((76/2)+1)}{2} = 38,5$$

La classe che contiene il valore mediano è la 300-400. **Me** è il valore centrale 350.

Classi di modalità	Frequenze assolute (ni)	Frequenze assolute cumulate
100-200	15	15
200-300	20	35
300-400	27	62
400-500	14	76
	76	



Esercizio

Considerando le seguenti frequenze assolute del livello di soddisfazione dei clienti di un servizio di vendita online, determinare il valore mediano.

x_i	Frequenze assolute	F
Per niente soddisfatto	28	
Poco soddisfatto	15	
Abbastanza soddisfatto	37	
Molto soddisfatto	21	
Non so	2	
	103	

Esercizio

Considerando le seguenti frequenze assolute del livello di soddisfazione dei clienti di un servizio di vendita online, determinare il valore mediano.

xi	Frequenze assolute	Frequenze assolute cumulate
Per niente soddisfatto	28	28
Poco soddisfatto	15	43
Abbastanza soddisfatto	37	80
Molto soddisfatto	21	101
Non so	2	103
	103	

$$(103+1)/2 = 52$$

Abb. soddisfatto



Metodo dell'interpolazione

Mediana

$$\mu_e = L_{inf} + c \frac{\frac{N}{2} - \sum F_{iMe}}{n_{Me}}$$

L_{inf} = limite inferiore della classe mediana

c = ampiezza della classe

F_{iMe} = frequenze cumulate della classe inferiore alla mediana

n_{me} = frequenza assoluta della classe mediana

Classi d'età	Frequenze assolute	Frequenze cumulate
0-20	22	22
20-40	45	67
40-65	80	147
65 ed oltre	40	187
	187	

$$= 40 + 25 * \frac{(187/2) - 67}{80}$$

Moda

Qual è il valore modale della distribuzione?

xi per classi	Frequenze % (pi)
20-30	36
30-40	57
40-50	51,5
50-60	47,5
60-70	34



Classe modale 30-40. E' possibile calcolare il valore centrale $(30+40)/2$ oppure utilizzare la formula:

$$\mu_0 = L_{\text{inf}} + c \frac{\Delta_{\text{inf}}}{\Delta_{\text{inf}} + \Delta_{\text{sup}}}$$

Esercizio

Qual è il valore modale della distribuzione?

xi per classi	Frequenze relative (fi)
35-45	0,179
45-55	0,329
55-70	0,451
70-85	0,381



In questo caso l'ampiezza delle classi è differente.
Nel caso di classi non sono **equiampie** bisogna calcolare le **densità di frequenza**.

Densità di frequenza

Rapporto tra le frequenze assolute (o relative) e l'ampiezza delle classi. Si può interpretare come il grado di addensamento delle frequenze nel corrispondente intervallo di valori.

$$d_i = \frac{n_i \text{ (o } f_i)}{h_i}$$

h_i = ampiezza delle classi (limite superiore - limite inferiore)

$$h_1 = 45 - 35 = 10$$

$$h_2 = 55 - 45 = 10$$

$$h_3 = 70 - 55 = 15$$

$$h_4 = 85 - 70 = 15$$

$$d_1 = 0,179 / 10 = 0,0179$$

$$d_2 = 0,329 / 10 = 0,0329$$

$$d_3 = 0,451 / 15 = 0,0300$$

$$d_4 = 0,381 / 15 = 0,0254$$

xi per classi	di
35-45	0,0179
45-55	0,0329
55-70	0,0300
70-85	0,0254

Intervallo modale è quello a cui corrisponde la più elevata densità di frequenza

Istogramma di densità

Se le classi sono equiampie non fa distinzione riportare la densità o la frequenza sull'asse verticale dell'istogramma.

Quando invece le classi non sono tutti uguali non sarebbe corretto disegnare l'altezza di ogni rettangolo in base alla frequenza. Infatti, un intervallo di valori più ampio tenderà a contenere un numero maggiore di frequenze.

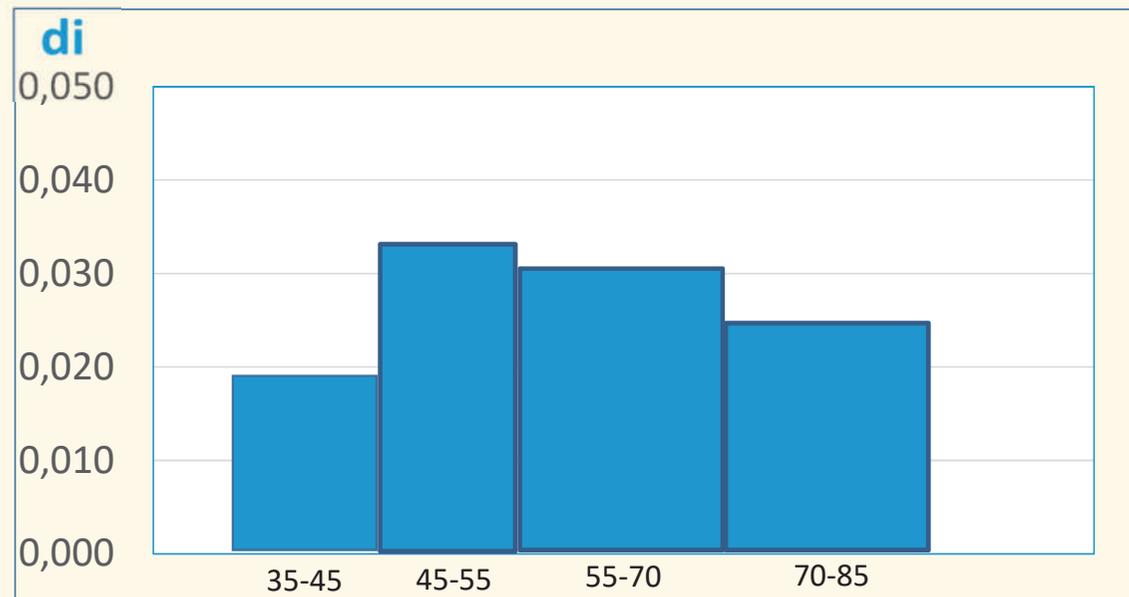
Negli **istogrammi di densità** (con classi di ampiezza diversa), l'altezza dei rettangoli indica la densità della classe, mentre la frequenza è rappresentata dall'area del rettangolo.

Frequenza = base x altezza

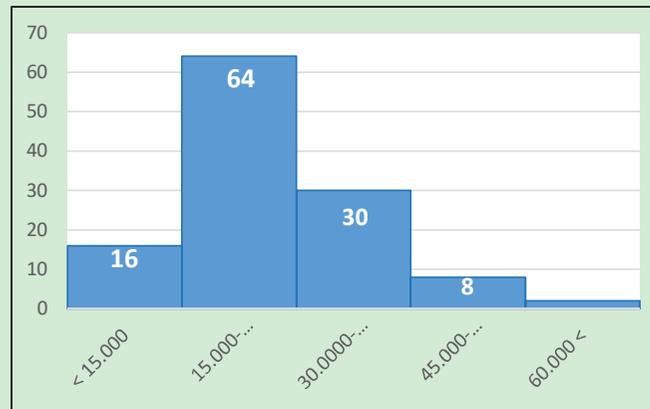
Altezza (ovvero densità) = $\frac{\text{frequenza}}{\text{base}}$

Istogramma di densità

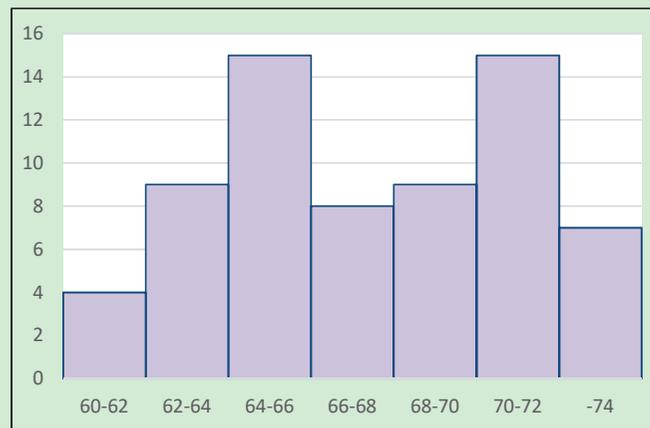
L'altezza dei rettangoli sarà determinata in modo che l'area del rettangolo sia pari alla frequenza della classe.
La somma delle aree sarà quindi la somma delle frequenze per ogni singola classe.



Tipi di distribuzione

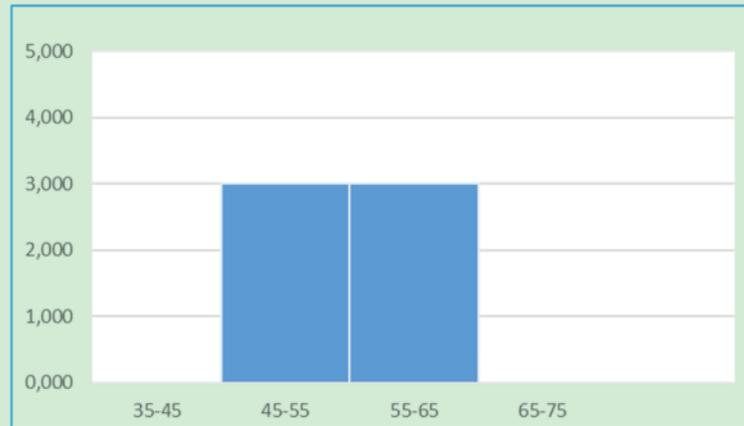


Unimodale

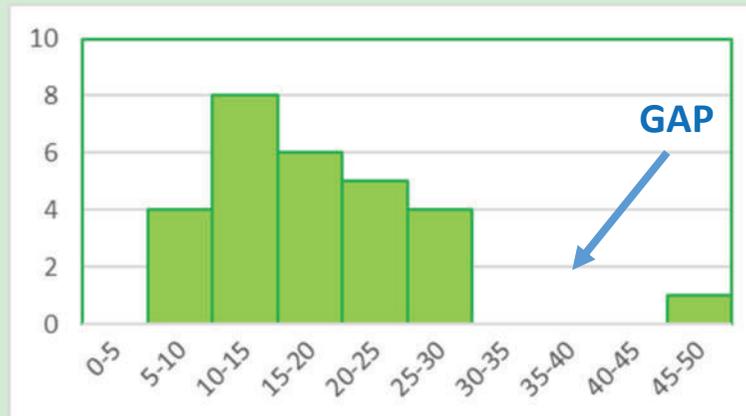


Bimodale

Tipi di distribuzione



Uniforme



Outlier

Riassunto

Carattere		Misura
Qualitativo	Sconnesso (nominale)	Moda
	Ordinabile	Moda, mediana, quartili
Quantitativo		Moda, mediana, quartili, media

Misura	Quando usarla	Caratteristiche
Media	Facile da calcolare	Indice poco robusto. Per distribuzioni con valori estremi non è appropriato
Media ponderata	Quando ogni osservazione ha un peso diverso	Indice poco robusto
Mediana	Per conoscere il punto centrale di una distribuzione o se questa è asimmetrica	Non è sensibile ai valori estremi
Moda	E' facile da usare per scale nominali	Misura grezza