- Medie Algebriche o Potenziate se la determinazione della *media* avviene utilizzando tutti i valori della distribuzione;
- Medie "lasche"(: Medie di Posizione e Moda) se la determinazione della *media* avviene utilizzando solo uno o alcuni valori della distribuzione.

LE MEDIE DI POSIZIONE

Definizione di MEDIANA: si definisce mediana quel termine che, nella successione ordinata dei valori, occupa la posizione centrale, ovvero quel termine che è preceduto e seguito dal 50% dei valori osservati.

CALCOLO DELLA MEDIANA DI X

I)X variabile discreta

a) dati semplici

-n numero dispari: $m_e = x_{(\frac{n+1}{2})}$

-n numero pari: $x_{(\frac{n}{2})} \leq m_e \leq x_{(\frac{n}{2}+1)}$

oppure la mediana può essere data dalla semisomma degli estremi.

b) dati ponderati: si procede in modo analogo al caso di dati semplici sostituendo ad n la somma delle frequenze N ed utilizzando le frequenze cumulate.

-N numero dispari: $m_e = x_{(\frac{N+1}{2})}$

-N numero pari: $x_{(\frac{N}{2})} \leq m_e \leq x_{(\frac{N}{2}+1)}$

Osservazione: la Mediana si può applicare anche a caratteri qualitativi, purchè su scala ordinale.

II)X variabile continua-valori divisi in classi: dapprima si individua la classe mediana secondo il procedimento al punto I) e poi, ipotizzando che all'interno della classe mediana i valori siano uniformemente distribuiti, si determina la mediana al suo interno nel seguente modo:

-se N è un numero dispari:

$$m_e = x_s + \frac{(x_{s+1} - x_s)}{f_s}\left(\frac{N+1}{2} - \sum_{i=1}^{s-1} f_i\right)$$

-se N è un numero pari:

$$m_e^{'} = x_s + \frac{(x_{s+1} - x_s)}{f_s}\left(\frac{N}{2} - \sum_{i=1}^{s-1} f_i\right)$$

$$m_e^{''} = x_s + \frac{(x_{s+1} - x_s)}{f_s}\left(\frac{N}{2} + 1 - \sum_{i=1}^{s-1} f_i\right)$$

Da cui $m_e = \dfrac{m_e^{'} + m_e^{''}}{2}$

ESERCIZIO

Dati relativi ad alcune regioni italiane (: variabile Y – n. di aziende di un certo settore economico):

Regioni	y
Piemonte	606
Lombardia	1974
Trentino A.A.	53
Friuli V.G.	132
Liguria	225
Emilia R.	556
Marche	338
Lazio	2035
Molise	69
Calabria	405
Sardegna	238

Determinare la Mediana di Y.

SOLUZIONI

y	*y ORD.*		
606	53	$y_{(1)}$	Trentino
1974	69		
53	132		
132	225		
225	238		
556	338	$y_{(6)}$	Marche
338	405		
2035	556		
69	606		
405	1974		
238	2035	$y_{(11)}$	Lazio

Innanzitutto bisogna ordinare i valori di Y, quindi si considera n (: n=11), che è un numero dispari:

Med(y)=$y_{\frac{n+1}{2}} = y_{\frac{11+1}{2}} = y_{(6)}$=338

ESERCIZIO

Sui seguenti valori determinare la Mediana di X:

x
5
15
25
35
45
55

I valori di X sono ordinati. Poiché n=6 è pari, allora:

$x_{n/2} \le Med(x) \le x_{n/2+1}$

$x_3 \le Med(x) \le x_4$

$25 \le Med(x) \le 35$

Ovvero, volendo approssimare: Med(x)=$\frac{25+35}{2}$=30

ESERCIZIO

Sui seguenti valori

x	f(x)
1	20
2	80
4	50
6	30
7	10

Determinare la Mediana di X.

SOLUZIONI

x	f(x)	F(x)
1	20	20
2	80	100
4	50	150
6	30	180
7	10	190
	190	

N=190 n.pari, allora:

$x_{N/2} \leq Med(x) \leq x_{N/2+1}$

$x_{95} \leq Med(x) \leq x_{96}$

Dalle frequenze cumulate F(x) si deduce: $x_{95}=2$ e $x_{96}=2$

Quindi Med(x)=2

ESERCIZIO

La durata in mesi x di una partita di batterie (1000) è risultata la seguente:

x	f(x)
0-1	50
1-3	180
3-6	300
6-12	320
12-20	150

Sui dati precedenti calcolare la Mediana di x.

SOLUZIONI

x	f(x)	F(x)
0-1	50	50
1-3	180	230
3-6	300	530
6-12	320	850
12-20	150	1000
	1000	

N=1000 n.pari, allora:

$x_{N/2} \leq Med(x) \leq x_{N/2+1}$

$x_{500} \leq Med(x) \leq x_{501}$

Dalle frequenze cumulate F(x) si deduce che entrambi i termini appartengono alla classe 3-6. Allora

$x_{500}=3+\frac{(6-3)}{300}(500-230)=5{,}7$

$x_{501}=3+\frac{(6-3)}{300}(501-230)=5{,}71$

5,7≤Med(x)≤5,71

Ovvero, volendo approssimare: Med(x)= $\frac{5{,}7+5{,}71}{2}$ =5,705

LE ***PROPRIETA'*** *DELLA MEDIANA*

I PROPRIETA': la mediana minimizza la somma degli scarti assoluti:
la quantità $D=\sum_{i=1}^{n}|x_i - A|$ è minima se A=Med(x).

Quindi la Mediana è il **Centro di grado 1**.

Ne consegue che la mediana consente la miglior allocazione di un servizio destinato a più utenti.

II PROPRIETA': la mediana non risulta influenzata da eventuali valori anomali presenti nella distribuzione.

Definizione di PERCENTILE: in una successione ordinata di valori si dice percentile il termine che ripartisce la distribuzione nella percentuale prefissata, ovvero in modo che la percentuale prefissata di dati non superi tale valore e che la percentuale complementare dei dati comprenda tutti gli elementi maggiori.

Variabile discreta (dati semplici): $x_{\%}=x_{\%(n)}$

Variabile discreta (dati ponderati): $x_{\%}=x_{\%(N)}$

Variabile continua in classi:

$$x_{\%} = x_{\%(N)} = x_s + \frac{(x_{s+1} - x_s)}{f_s}\left[\%(N) - \sum_{i=1}^{s-1} f_i\right]$$

ESERCIZIO

Sui seguenti valori

x	f(x)
1	20
2	80
4	50
6	30
7	10

Determinare il 70° Percentile di X.

SOLUZIONI

x	f(x)	F(x)
1	20	20
2	80	100
4	50	150
6	30	180
7	10	190
	190	

-Calcolo del 70° percentile $\mathbf{x_{70\%}}$:

Il 70% di N=190 è 133, quindi $x_{70\%}=x_{(133)}=4$ dalle frequenze cumulate.

ESERCIZIO

Dati i seguenti valori ***x*** e le rispettive frequenze ***f(x)***

x	f(x)
1	40
3	80
5	60
7	20

Calcolare il 10° Percentile.

SOLUZIONI

x	f(x)	F(x)
1	40	40
3	80	120
5	60	180
7	20	200
	200	

-Calcolo del 10° percentile $\mathbf{x_{10\%}}$:

Il 10% di N=200 è 20, quindi $x_{10\%}=x_{(20)}=1$ dalle frequenze cumulate.

ESERCIZIO

Sui seguenti valori

x	f(x)
x≤ 2	5
2<x≤ 4	60
4<x≤ 6	355
6<x≤ 8	70
x>8	10

Calcolare il 90° Percentile della variabile X.

SOLUZIONI

x	f(x)	F(x)
x≤2	5	5
2<x≤4	60	65
4<x≤6	355	420
6<x≤8	70	490
x>8	10	500

Calcolo del 90° percentile: il 90% di N=500 è 450, quindi si deduce dalle frequenze cumulate che $x_{90\%}=x_{(450)}$ appartiene alla **classe 6-8**.
Allora:

$$x_{450}=6+\frac{(8-6)}{70}(450-420)=6{,}857$$

ESERCIZIO

La durata in mesi x di una partita di batterie (1000) è risultata la seguente:

x	f(x)
0-1	50
1-3	180
3-6	300
6-12	320
12-20	150

Sui dati precedenti calcolare il 75° Percentile di x.

SOLUZIONI

x	f(x)	F(x)
0-1	50	50
1-3	180	230
3-6	300	530
6-12	320	850
12-20	150	1000
	1000	

Il 75% di N=1000 è 750, quindi si deduce dalle frequenze cumulate che $x_{75\%}=x_{(750)}$ appartiene alla **classe 6-12**.
Allora:

$$x_{750}=6+\frac{(12-6)}{320}(750-530)=10,125$$

Definizione di QUARTILE: in una successione ordinata di valori si dicono quartili quei termini che ripartiscono la distribuzione in 4 parti uguali:

$$Q_1 = x_{1/4} = x_{25\%}$$

$$Q_2 = x_{1/2} = x_{50\%} \equiv \text{Mediana}(x)$$

$$Q_3 = x_{3/4} = x_{75\%}$$

$Q_4 = x_{4/4} = x_{100\%}$ **(è l'ultimo termine della distribuzione)**

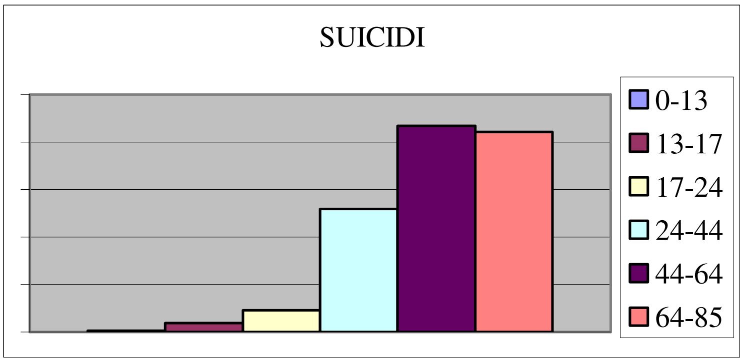
ESERCIZIO

Nella tabella che segue vengono indicate le distribuzioni per classi di età dei suicidi e dei tentativi di suicidio verificatisi nel 1987 nell'ambito della popolazione femminile (fonte ISTAT 1988). Confrontare i Quartili delle due distribuzioni.

età X	SUICIDI f(x)	TENTATIVI di SUICIDIO f'(x)	F(x)	F'(x)
0-13	3	5	3	5
13-17	19	101	22	106
17-24	46	236	68	342
24-44	259	526	327	868
44-64	434	321	761	1189
64-85	421	142	1182	1331
TOTALI	1182	1331		

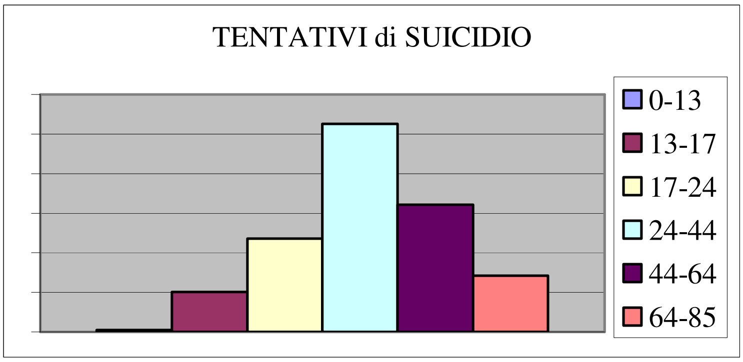
Per la distribuzione dei SUICIDI:

Q_1=42; Q_2=56; Q_3=70 Q_4=85



Per la distribuzione dei TENTATIVI di SUICIDIO:

Q_1=24; Q_2=36; Q_3=52; Q_4=85



Definizione di MODA: data la distribuzione di frequenze relativa ad un certo carattere, si definisce moda la modalità di quel carattere che si presenta con la massima frequenza.

ESERCIZIO

Sui seguenti valori

x	f(x)
1	20
2	80
4	50
6	30
7	10

Determinare la Moda.

SOLUZIONI

Moda(x)=2, termine della distribuzione cui corrisponde la massima frequenza.

ESERCIZIO

Sui seguenti valori

x	f(x)
$0<x\leq 2$	5
$2<x\leq 4$	60
$4<x\leq 6$	355
$6<x\leq 8$	70
$8<x\leq 10$	10

Calcolare la Moda.

SOLUZIONI

MODA(x): **Classe modale(x)**= classe **4-6** cui corrisponde la massima frequenza, pari a 355 (in questa situazione non è necessario il calcolo delle *Densità di frequenza* poiché le classi sono tutte di ampiezza uguale e pari a 2).

FUNZIONE DI DANNO

$$D = \sum_{i=1}^{n} \left| x_i - A \right|^{s}$$

Si definisce ***centro di grado s*** la quantità ***A*** che minimizza la funzione di danno D.

Si dimostra che:

la Moda è il **Centro di grado 0**.

la Mediana è il **Centro di grado 1**.

la Media Aritmetica è il **Centro di grado 2**.