SINTESI DELLE DISTRIBUZIONI UNIVARIATE
LE MEDIE

Rappresentazione dei dati
- **Tabellare**
- **Grafica**
- **Sintetica (indici)**
- **Analitica**

Sostituire la distribuzione (successione) con un unico valore allo scopo di mettere in evidenza un particolare aspetto della distribuzione stessa

Indici di
- **Posizione**
- **Variabilità**
- **Forma**

CARATTERI QUANTITATIVI DISCRETI

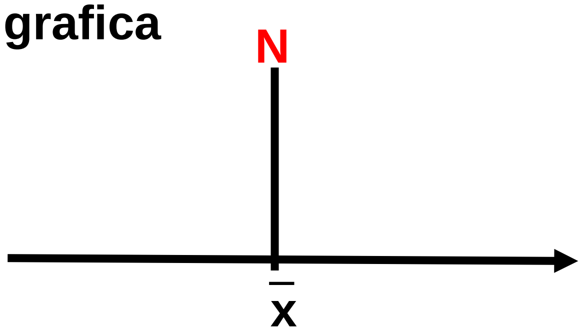
Se il carattere è quantitativo discreto (scala di intervallo o di rapporto), le modalità della distribuzione possono essere rappresentata su una retta come un insieme di punti, a ciascuno dei quali è applicato un "peso" proporzionale alla frequenza con cui la modalità corrispondente si presenta nel collettivo di N unità.

Indice di posizione: il valore che esprime sinteticamente la posizione della distribuzione sull'asse reale

Caso banale: tutte le unità presentano la stessa modalità → $X_j = \bar{x}\ \forall j$

Rappres. tabellare **Rappres. grafica**

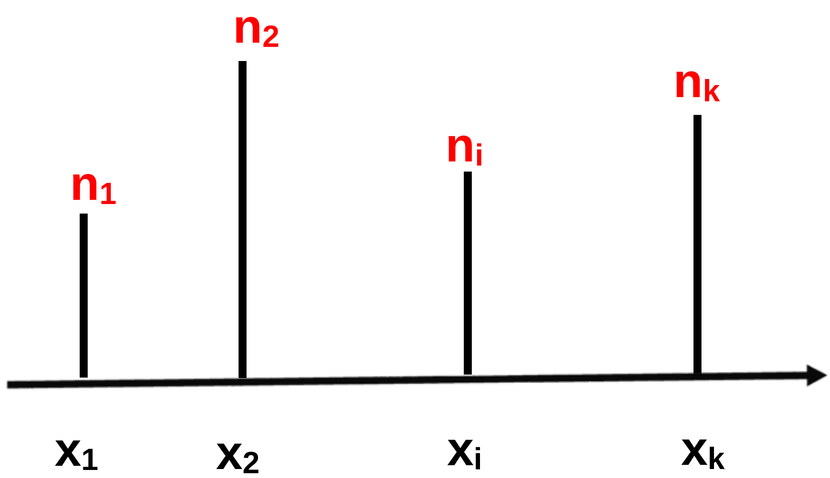
Carat-tere X	Frequenza
$\bar{x}$	N



In questo caso il valore $\bar{x}$ rappresenta esattamente la posizione della distribuzione sull'asse.

In generale

modalità del carattere X	*frequenza*
x_1	n_1
x_2	n_2
…	…
x_i	n_i
…	…
x_k	n_k
	N



Come scegliere il valore $\bar{x}$ chiamato a rappresentare la posizione della distribuzione sull'asse.

CONDIZIONE DI CAUCHY (internalità): $x_1 \leq \bar{x} \leq x_k$

Condizione concettualmente rilevante ma operativamente poco utile

Un approccio possibile

Medie dedotte da condizione di equivalenza

Si abbia una successione $(x_1, x_2, \ldots, x_j, \ldots, x_N)$ di modalità di un carattere quantitativo osservato su N unità statistiche e si consideri una funzione *$f(x_1, x_2, \ldots, x_j, \ldots, x_N)$, che rappresenti una caratteristica rilevante della successione.*
Si dice media della successione $(x_j)_{j=1,2,\ldots,N}$ rispetto ad $f(\cdot)$ quel valore $\bar{x}$ **tale che**

$$f(x_1, x_2, \ldots, x_j, \ldots, x_N) = f(\underbrace{\bar{x}, \bar{x}, \ldots, \bar{x}}_{\text{N volte}})$$

In altri termini: si sostituiscono I valori osservati con un unico valore costante, destinato a rappresentarli tutti, con la condizione che la sostituzione non alteri una caratteristica della successione che si assume come "invariante"

La funzione *$f(\cdot)$* può essere scelta in un qualsiasi modo, purché soddisfi la condizione di internalità

ATTN Si avranno tanti tipi di media quanti sono i modi di scegliere la funzione *$f(\cdot)$.*

La scelta dipende

- **dalla natura del carattere**
- **dal tipo di problema che si vuol risolvere sostituendo tutti i valori con una costante (dal tipo di caratteristica che si vuol lasciare invariata)**

QUALCHE ESEMPIO

Esempio 1 ***$(x_j)_{j=1,\dots,N}$ retribuzioni di N addetti in una impresa***

Salario medio?

$$f(x_1, x_2, \dots, x_j, \dots, x_N) = f(\bar{x}, \bar{x}, \dots\dots, \bar{x})$$

monte salari

invariante $\Sigma_j\, x_j \quad = \quad \Sigma_j\, \bar{x} \quad = N\ \bar{x}$

carattere additivo
media aritmetica

$$\bar{x} = \frac{\sum_j x_j}{N}$$

****°°°°°****

Esempio2 ***$(x_j)_{j=1,\dots,N}$ lati di N appezzamenti di terreno quadrati***

Lato medio?

$$f(x_1, x_2, \dots, x_j, \dots, x_N) = f(\bar{x}, \bar{x}, \dots\dots, \bar{x})$$

superficie totale

invariante $\sum_j x_j^2 = \sum_j (\bar{x})^2 = N\ (\bar{x})^2$

car. additivo nei quadrati
media quadratica

$$\bar{x} = \sqrt{\frac{\sum_j x_j^2}{N}}$$

****°°°°°****

Esempio 3 *$(x_j)_{j=1,\dots,N}$ velocità a cui si percorre N volte un tratto di strada della stessa lunghezza*

Supponiamo che il tratto di strada sia di 60km. Lo si percorre all'andata a 120 km/h e al ritorno a 30 km/h. Qual è la velocità media tenuta nel percorso A/R?

<u>1° modo di affrontare il problema</u>
Media aritmetica → (120+30)/2 = 75 km/h

Per completare i 60+60=120 km del percorso A/R a 75 km/h impiego 120/75= 1,6 ore, ovvero 1 ora e 36 minuti

<u>2° modo di affrontare il problema</u>
All'andata, viaggiando a di 120 km/h, impiego 60/120=0,5 ore, ovvero 30 minuti. Al ritorno, con velocità di 30 km/h, per fare 60 km impiego 2 ore. Quindi per l'intero tratto occorrono 2 ore e 30 minuti.

Se voglio la velocità media intesa come quella che mantenuta costantemente nelle N prove mi lascia inalterato il tempo di percorrenza complessivo, il modo corretto di affrontare il problema è il 2°.

$$\frac{60}{120}+\frac{60}{30}=\frac{60}{\bar{x}}+\frac{60}{\bar{x}} \rightarrow \frac{1}{120}+\frac{1}{30}=\frac{1}{\bar{x}}+\frac{1}{\bar{x}} \text{ da cui } \bar{x}=\frac{2}{\frac{1}{120}+\frac{1}{30}} = 48 \text{ km/h}$$

Infatti, viaggiando ad una velocità di 48 km/h, per percorrere 60+60=120 km occorrono 120/48=2,5 ore

$$f(x_1, x_2, \dots, x_j, \dots, x_N) = f(\bar{x}, \bar{x}, \dots\dots, \bar{x})$$

somma dei reciproci
invariante

$$\sum_j \frac{1}{x_j} = \sum_j \frac{1}{\bar{x}} \qquad \textit{da cui} \quad \bar{x}=\frac{N}{\sum_j \frac{1}{x_j}}$$

carattere additivo nei reciproci
media armonica

Esempio 3bis ***$(x_j)_{j=1,\dots,N}$*** ***durata di una confezione di fette biscottate***

Supponiamo 4 famiglie che rispondono alla domanda "quanti giorni vi dura una confezione?". Le risposte siano

Fam	*durata gg*
A	*6*
B	*6*
C	*4*
D	*8*

Durata media (6+6+4+8)/4=6gg
Quindi, ciascuna famiglia consuma in media 360/6=60 confezioni all'anno e una produzione di 60*4 = 240 confezioni soddisfa le esigenze di consumo della collettività delle 4 famiglie

MA: il consumo annuo delle 4 famiglie è

Fam	*conf all'anno*
A	*360/6=60*
B	*360/6=60*
C	*360/4=90*
D	*360/8=45*

Totale	**255**

monte salari

?

Consumo **invariante**

$$\frac{360}{6}+\frac{360}{6}+\frac{360}{4}+\frac{360}{8}=\frac{360}{\bar{x}}+\frac{360}{\bar{x}}+\frac{360}{\bar{x}}+\frac{360}{\bar{x}}$$

$$\frac{1}{6}+\frac{1}{6}+\frac{1}{4}+\frac{1}{8}=\frac{4}{\bar{x}} \quad da\ cui \quad \bar{x}=\frac{4}{\frac{1}{6}+\frac{1}{6}+\frac{1}{4}+\frac{1}{8}}=5,647$$

carattere additivo nei reciproci

media armonica

Esempio 4 ***$(x_j)_{j=1,\dots,N}$ tassi di interesse applicati in N anni successivi.***

Investo un capitale C per N anni a quei tassi. Qual è il tasso medio, ovvero quel tasso che se applicato costantemente per tutti gli anni mi darebbe lo stesso capitale finale?

Anno	*tasso %*	*fattore di capitalizzazione*
1°	*6%*	*1,06*
2°	*4%*	*1,04*
3°	*2%*	*1,02*

Se investo un capitale C, alla fine del terzo anno avrò

C(1,06)(1,04)(1,02) = montante finale

Il fattore di capitalizzazione medio è quello che applicato costantemente per i 3 anni lascia invariato il montante

$$(1,06)(1,04)(1,02) = (\bar{x})(\bar{x})(\bar{x}) = (\bar{x})^3$$

$$da\ cui\ \ \bar{x} = \sqrt[3]{(1,06)(1,04)(1,02)} \approx 1,0399$$

Il tasso medio dell'investimento è di 3,99%, leggermente inferiore al tasso medio aritmetico (6+4+2)/3=4.

$$\prod_{j=1}^{N} x_j = \prod_{j=1}^{N} \bar{x} = \bar{x}^N \quad da\ cui\ \ \bar{x} = \sqrt[N]{\prod_{j=1}^{N} x_j}$$

car. moltiplicativo

media geometrica

Esempio 5 media aritmetica ponderata

Nella media aritmetica

$$\bar{x} = \sum_j x_j \frac{1}{N}$$

ogni unità vale 1/N (ha lo stesso peso). Talvolta può essere opportuno dare alle varie unità peso differente.

Esempio siano $(x_j)_{j=1,\dots,N}$ i quozienti di natalità dei comuni toscani. Ogni modalità è espressione del rapporto n_j/p_j tra nati e popolazione del comune j-esimo. Si vuol calcolare il quoziente medio (regionale).

Una ragionevole condizione di invarianza è: sostituiamo i tassi di natalità dei vari comuni con un valore costante che lasci invariato il numero totale di nati della regione.

$$\sum_j x_j\, p_j = \sum_j n_j = \sum_j \bar{x}\, p_j \quad da\ cui \quad \bar{x} = \frac{\sum_j x_j\, p_j}{\sum_j p_j}$$

Media ponderata

Un caso notevole: Indice sintetico dei prezzi

siano $({}_0x_{tj})_{j=1,\dots,N}$ gli indici elementari del prezzo di N beni, espressione, ciascuno, del rapporto p_{tj}/p_{0j} tra il prezzo del bene j al tempo t e il prezzo dello stesso bene al tempo 0. Siano q_0 e q_t la quantità ai tempi 0 e t.

Sostituiamo tutti i valori con uno costante che lasci invariata la quantità acquistata al tempo 0

$$\sum_j {}_0x_{tj}\,(p_{0j}q_{0j}) = \sum_j p_{tj}\, q_{0j} = \sum_j \bar{x}\, p_{0j}q_{0j} \quad da\ cui$$

$$\bar{x} = \frac{\sum_j {}_0x_{tj}\, p_{0j}q_{0j}}{\sum_j p_{0j}\, q_{0j}} = \frac{\sum_j p_{tj}\, q_{0j}}{\sum_j p_{0j}\, q_{0j}}$$

Indice dei prezzi di Laspeyres

Riepilogo delle Formule

Successione $(x_j)_{j=1,2,\dots,N}$	Media	Distr. frequenza $(x_i,\ n_i)_{i=1,2,\dots,k}$
$\frac{\sum_j x_j}{N}$	aritmetica	$\frac{\sum_i x_i n_i}{\sum_i n_i}$
$\sqrt{\frac{\sum_j x_j^2}{N}}$	quadratica	$\sqrt{\frac{\sum_i x_i^2 n_i}{\sum_i n_i}}$
$\frac{N}{\sum_j \frac{1}{x_j}}$	armonica	$\frac{\sum_i n_i}{\sum_i \frac{n_i}{x_i}}$
$\sqrt[N]{\prod_{j=1}^{N} x_j}$	geometrica	$\sqrt[\sum_i n_i]{\prod_{i=1}^{k} x_i n_i}$
$\frac{\sum_j x_j p_j}{\sum_j p_j}$	ponderata	$\frac{\sum_i x_i n_i p_i}{\sum_i n_i p_i}$

E nel caso di distribuzioni di frequenza

$(x_i – x_{i+1},\ n_i)_{i=1,2,\dots,k}$

con modalità raggruppate in classi?

Si sostituisce alla classe $x_i – x_{i+1}$, il valore centrale di classe $c_i=(x_i+x_{i+1})/2$ e ci si riconduce al caso non raggruppato in classi sostituendo c_i al posto di x_i

PROPRIETA' DELLA MEDIA ARITMETICA

La media aritmetica è la media più nota e la più usata (talvolta abusata)

Indichiamo con M(X) l'operatore media aritmetica del carattere X. Esso gode delle seguenti proprietà

Internalità $x_{min} \leq M(X) \leq x_{max}$

Omogeneità $M(aX) = a\, M(X)$

Traslatività $M(a+X) = a + M(X)$

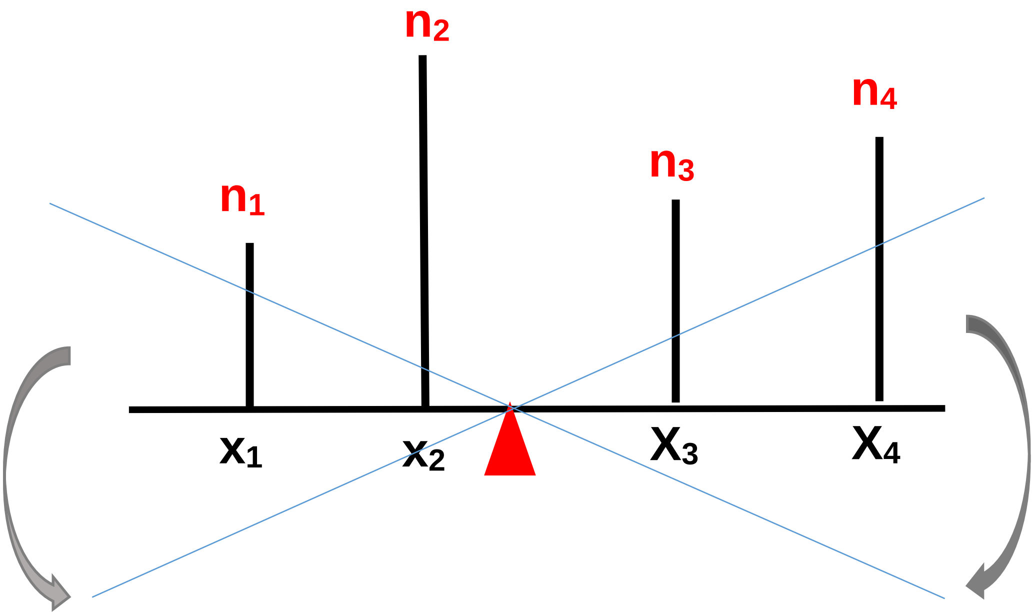
Associatività

$\underbrace{x_1, x_2, \ldots, x_j,}_{\text{gruppo A}} \quad \underbrace{x_{j+1}, x_{j+2}, \ldots, x_N}_{\text{gruppo B}}$

$$M(X) = \left[M(X \in A) \times j + M(X \in B) \times (N-j)\right] / N$$

La somma degli scostamenti dalla media aritmetica è pari a zero (baricentro)

$$\sum_i [x_i - M(X)]\, n_i = 0$$



La somma dei quadrati degli scostamenti dalla media aritmetica è un minimo

$$\sum_i [x_i - M(X)]^2\, n_i = min$$

La media aritmetica è il punto più vicino a tutti i valori della distribuzione secondo la definizione di distanza espressa come somma degli scostamenti al quadrato

ALTRI INDICI DI POSIZIONE

MODA: la modalità che si presenta con la frequenza maggiore
(determinabile anche per caratteri misurati su scala nominale)

Occupati in Provincia di Firenze per sezioni di attività economica
Fonte: Istat, Censimento della popolazione 2011

Sezioni attività economica	occupati	
Agricoltura, silvicoltura e pesca	12.122	
Industria	111.224	
Servizi	297.707	MODA
totale	421.053	

MEDIANA: valore che nella graduatoria ordinata dei valori occupa la posizione centrale (il numero di valori che lo precedono è uguale al numero di valori che lo seguono)
(determinabile anche per caratteri misurati su scala nominale)

Individui classificati per titolo di studio ed età

titolo di studio	Età			Totale	Freq cumul	
	25-35	35-55	55-65			
Nessun titolo	10	20	50	80	80	
Licenza elementare o media	140	130	100	370	450	
Diploma	100	220	100	420	870	Mediana
Laurea	50	50	30	130	1000	
Totale	300	420	280	1000		

tit di studio mediano → modalità dell'unità statistica che sta tra 500-esimo e 501-esimo posto in graduatoria

età mediana?

Individui classificati per titolo di studio ed età

titolo di studio	Età			Totale
	25-35	35-55	55-65	
Nessun titolo	10	20	50	80
Licenza elementare o media	140	130	100	370
Diploma	100	220	100	420
Laurea	50	50	30	130
Totale	300	420	280	1000
Freq cumul	300	720	1000	

Età mediana → età dell'unità statistica che sta tra 500-esimo e 501-esimo posto in graduatoria

È certamente nella classe 35-55 ma dove?

Ipotesi: distribuzione uniforme dei 420 individui all'interno della classe 35-55.

Mediana = 35 + 20×(200/420) = 35 + 9,52 = 44,52

p-esimo PERCENTILE**: valore che nella graduatoria ordinata dei valori occupa una posizione tale che il p% dei valori è più piccolo e il (100-p)% dei valori è più grande**