Statistica 1 - Teoria

Lezione 1 - 4/10/21

Intro : cos'é la Statistica

: disciplina che tratta, con linguaggio matematico, principi e metodi che sono d ausilio all'indagine sperimentale, condotta da altre discipline.

Interviene, in modo astratto e generale, rispetto ai risultati sperimentali (per ottenerli/interpretarli), a prescindere dai contenuti delle altre discipline.

La variabilitá accidentale per fenomeni ripetibili

La statistica studia fenomeni aleatori ripetibili

possono manifestarsi più volte rimanendo fissi nei loro aspetti essenziali

le manifestazioni risultano però diverse di volta in volta e in modo imprevedibile, in base agli aspetti fissi e noti

presentano variabilità accidentale

Si distinguono due classi di fenomeni aleatori ripetibili:

A Fenomeni caratterizzati da ripetitibilità virtuale

: possono manifestarsi idealmente infinite volte, in base ad un meccanismo che li genera, che chiamiamo circostanze, il quale nonostante sia costante (per quanto si sappia) produce risultati variabili.

Esempio: lancio della moneta:

1. se la si lancia a mano, l'esito sarà circa 50 e 50; le circostanze tuttavia non sono davvero costanti, in quanto non è possibie garantire uguali lanci a mano succesivi.

2. assumiamo allora di lanciare tramite una macchina: il lancio ora sarà costante, la varibilità acidentale ridotta, ma non annullata, in quanto l'esito del fenomento dipenderàm anche da altri elementi, come densità dell'aria e irregolarità del piano.

3. ipotizziamo allora la macchina sottovuoto e il piano perfettamente liscio: elimineremo altra variabilità, ma comunque non del tutto. Ss

Morale : origine della variabilità accidentale : le circostanze si assumono costanti quando in realtà non lo sono realmente

Eliminare variabilità accidentale $\Longleftrightarrow$ descrivere e conoscere con estrema precisione il fenomeno

a volte può essere inutile e/o molto costoso

Due approcci:

1) Punto di vista deterministico

Si cerca di eliminare la variabilità (necessario in alcuni problemi es. sicurezza centrali nucleari)

2) Punto di vista statistico

Si accetta la variabilità accidentale e si cerca di creare un modello di previsione delle future manifestazioni che tengono conto delle circostanze che si conoscono e della potenziale variabilità

Ibrido: in molti casi reali, se la variabilità è eccessiva la si cerca di ridurla ad un livello accettabile eliminandone alcune cause , poi si crea il modello.

B Fenomeni caratterizzati da ripetibilità attuale

: fenomeni aleatori ripetibile le cui realizzazioni si sono già manifestate all'istante di inizio dello studio

es. demografia : se studio caratteristiche antropometriche di un gruppo di individui, esse si sono già "manifestate", in quanto si trovano come dati anagrafici su Istat ecc.

In questo caso, a generare variabilità accidentale è l'incompletezza delle specificazioni : le specificazioni che non sono conosciute allo statistico ma che influenzano la manifestazione del fenomeno

reciproco di circostanze

es. fenomeno da studiare = reddito di operai : dipende da molti fattori; magari quelli a disposizione dello statistico non sono sufficienti a spiegare la variabilità del reddito (=manifestazione del fenomeno)

=> si ha variabilità accidentale (concettualmente uguale a ripetibilità virtuale)

Anche qui due approcci:

a) approfondire tantissimo lo studio per avere tutte le informazioni necessarie a spiegare la variabilità (molto costoso / impossibile)

b) accettare la variabilità accidentale

Quindi: la statistica tenta di scoprire la componente strutturale di un fenomeno ripetibile, accettando la variabilità accidentale causata dall'incompletezza di informazioni/conoscenza.

Tipologie di variabili/caratteri/dati

Qualitativo = esprime una caratteristica del fenomeno qualitativamente, attraverso sostantivi, aggettivi (in generale = a parole)

- **Ordinabile** = le modalità del carattere possono essere espresse in scala ordinale, secondo un criterio "naturale", o comunque universalmente riconosciuto.
  es: titolo di studio, piramide aziendale,...
- **Sconnesso** = le modalità del carattere sono esprimibili su scala nominale, o comunque non secondo criteri di ordine "naturale" o universalmente riconosciuto.
  es: comune di residenza, colori, sesso, punti cardinali,...

Si usano due scale per i caratteri qualitativi:

① Scala **nominale** : le modalità sono aggettivi (/sostantivi/avverbi) NON ORDINABILI (qual. sconnessi)
(non posso dire se $x_1 \gtrless x_2$, ma solo se $x_1 = x_2$ o $x_1 \neq x_2$)

② Scala **ordinale** : le modalità sono " ORDINABILI (qual. ordinabili)
(posso dire se $x_1 \gtrless x_2$)

Quantitativo = modalità espresse da numeri reali, nella forma di misurazioni o conteggi. Descrivono una proprietà oggettiva dell'unità statistica.

- **Continuo** : definito in un insieme di numeri reali
  es: peso, statura, tempo, ...
- **Discreto** = definito in un insieme di numeri interi

Per definizione vi è un **ordinamento** delle modalità.

Con i caratteri quantitativi, è possibile ha senso effettuare le seguenti operazioni:

Siano x_i = modalità del carattere X; allora hanno senso le operazioni:

$x_1 = x_2$; $x_1 \neq x_2$; $x_1 \gtrless x_2$; distanza $= |x_2 - x_1| = |x_1 - x_2|$

Si possono adottare, con i caratteri quantitativi, due diverse scale:

① **Scala a intervalli** (intervallare) : • lo zero è convenzionale, arbitrario

es: temperatura in scala Celsius, Farenheit : lo 0 è diverso

Non ha senso valutare gli incrementi percentuale : x_2/x_1

② **Scala di rapporti** (proporzionale) : • lo zero è oggettivo, ed indica assenza di quantità
• il valore della modalità è sempre ≥ 0

es: grandezze fisiche, valuta*, .. Infatti : 0 kg = 0 libbre ; 0 € = 0 $

Ha senso valutare la variazione percentuale x_2/x_1

*NB: la valuta diventa negativa, ma in realtà se le mie spese superano le entrate è possibile perché avrò preso in prestito

Lezione 3 - 7/10/21

Spoglio dei dati

: organizzazione dei dati in tabelle sintetiche.

Consideriamo per ora una variabile alla volta : statistica UNIVARIATA.

Tabella-tipo riportante la DISTRIBUZIONE DI FREQUENZE :

x_i	n_i	$f_i = n_i/n$
x_1	n_1	$n_1/n = f_1$
x_2	n_2	$n_2/n = f_2$
...	...	...
x_m	n_m	$n_m/n = f_m$
Totale	**n**	**1**

x_i : MODALITÀ assunte dal carattere X = manifestazioni della variabile statistica X

n_i : FREQUENZE ASSOLUTE per ogni modalità x_i; conta quante volte appare la modalità i-esima

f_i : FREQUENZE RELATIVE per ogni modalità x_i; misura relativa della frequenza di manifestazione della modalità i-esima

Freq. assolute n_i $\begin{cases} n_i \geq 0 \\ \sum_{i=1}^{m} n_i = n \end{cases}$ = numero totale di osservazioni

Freq. relative $f_i = \frac{n_i}{n}$ $\begin{cases} 0 \leq f_i \leq 1 \\ \sum_{i=1}^{m} f_i = 1 \end{cases}$

Freq. relative PERCENTUALI $p_i\% = f_i \cdot 100$ $\begin{cases} 0 \leq p_i \leq 100 \\ \sum_{i=1}^{m} p_i = 100 \end{cases}$

NB : a volte una variabile qualitativa viene chiamata MUTABILE.

ES.

X = categoria

(qualitativa ordinabile => in tabella si mettono rispettando l'ordine*)

x_i	n_i	f_i	$p_i\%$
Dettagliante	6	0,3	30
Grande distr.	4	0,2	20
Grossista	10	0,5	50
	20	1	100

ES.

X = unità vendute a singolo cliente

↳ Quantitativa discreta

x_i	n_i	f_i	$p_i\%$
3	3	3/20	;
4	8	8/20	
5	9	9/20	
	20	1	100

* Per le qualitative sconnesse, si mettono in tabella "a caso", in un ordine arbitrario

Per le quantitative continue, bisogna costruire delle CLASSI all'interno dell'intervallo $x_{min} \vdash x_{max}$

di AMPIEZZA VARIABILE o FISSA

I_i	n_i	$f_i = n_i/n$	N_i *	d_i	a_i
$x_0 \vdash\!\dashv x_1$	n_1	f_1	$N_1 = n_1$	d_1	a_1
$x_1 (\dashv] x_2$ escluso incluso	n_2	f_2	$N_2 = n_1 + n_2$	d_2	a_2
⋮	...	...	...	...	⋮
$x_{i-1} \dashv x_i$	n_i	f_i	$N_i = \sum_{j=1}^{i} n_j$	d_i	a_i
⋮	...	...	...	...	⋮
$x_{m-1} \dashv x_m$	n_m	f_m	$N_m = n$	d_m	a_m
	n	1			

* N_i = Frequenze CUMULATE ASSOLUTE (prossima pagina)

DENSITÀ DI FREQUENZA $d_i = \frac{n_i}{a_i}$

AMPIEZZA della CLASSE $a_i = x_i - x_{i-1}$

DENSITÀ DI FREQUENZA RELATIVA $d_i^* = \frac{f_i}{a_i}$ (meglio usare d_i)

NB: all'esame sarà richiesto esplicitare la formula che si usa (es. $d_3 = \frac{n_3}{a_3}$)

A volte si usa anche il VALORE CENTRALE DELLA CLASSE $x_{i_c} = \frac{x_i - x_{i-1}}{2}$

Es.

X = Fatturato

Ampiezza costante delle classi

I_i	n_i	a_i	d_i	x_{ic}
$0 \vdash\!\dashv 10$	4	10	4/10	5
$10 \dashv 20$	8	10	8/10	15
$20 \dashv 30$	4	10	4/10	25
$30 \dashv 40$	3	10	3/10	35
$40 \dashv 50$	1	10	1/10	45
	20			

Ampiezza variabile delle classi

I_i	n_i	a_i	d_i
$0 \vdash\!\dashv 10$	4	10	4/10 = 0,4
$10 \dashv 20$	8	10	8/10 = 0,8
$20 \dashv 30$	4	10	4/10 = 0,4
$30 \dashv 50$	4	20	4/20 = 0,2
	20		

NB! Notare come nel caso:
- di ampiezza costante, le d_i non danno alcuna informazione in più
- di ampiezza variabile, le d_i sono molto rilevanti!

Lezione 4 - 11/10/21

Frequenze cumulate

(i) Assolute N_i : $\{ n_i ; i=1, \dots, k \}$

$N_1 = n_1$, $N_2 = n_1 + n_2$, ... , $N_k = n_1 + n_2 + \dots + n_k = n$

(ii) Relative F_i : $\{ x_i ; i = 1, \dots, k \}$

$F_1 = f_1$, $F_2 = f_1 + f_2$, ... , $F_k = f_1 + f_2 + \dots + f_k = 1$

NB: usiamo sempre 3 decimali

In formule:

a) Formula compatta : $N_i = \sum_{j=1}^{i} n_j$ con $j = 1, \dots, i, \dots, k$

indica fino a dove devo sommare => limite superiore della Σ

b) Formula ricorsiva : $N_i = N_{i-1} + n_i$ per $i = 2, \dots, k$

Esempio

I_i	n_i	N_i	f_i	F_i
0 ⊣ 10	4	4	0,2	0,2
10 ⊣ 20	8	12	0,4	0,6
20 ⊣ 30	4	16	0,2	0,8
30 ⊣ 40	3	19	0,15	0,95
40 ⊣ 50	1	20	0,05	1
	20		1	

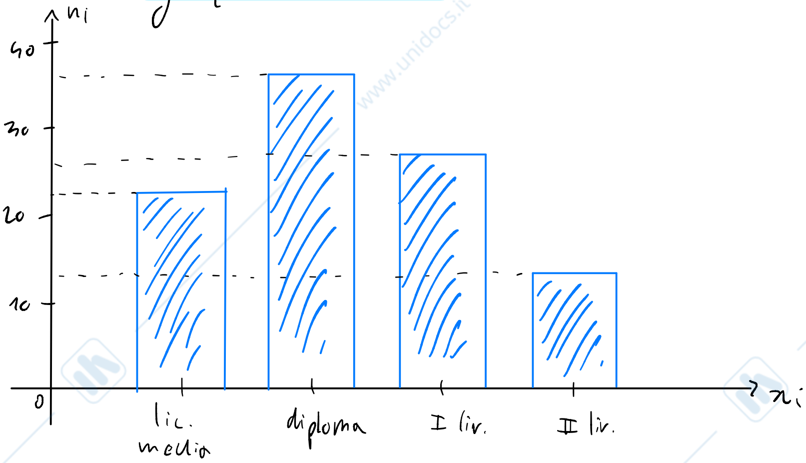
Grafici

Abbiniamo ad ogni tipo di carattere una sola rappresentazione grafica.

NB Consideriamo le frequenze assolute per costruire il grafico; per le relative basta riscalarlo.

- Qualitativo ordinabile : grafico a barre

x_i	n_i
lic. media	23
diploma	36
laurea 1° liv.	27
laurea 2° liv.	14

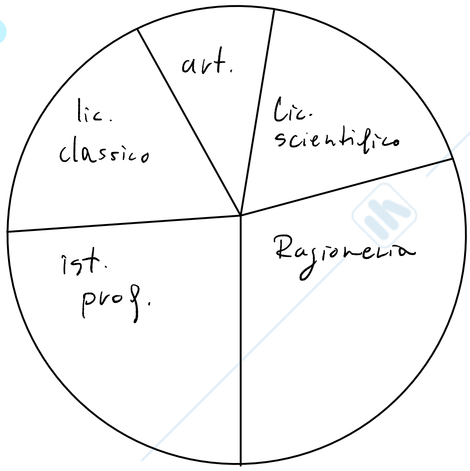


le x_i sono equidistanti, quindi non posso quantificare la distanza, il ché mi va bene data la natura qualitativa del carattere.

- Il grafico a barre preserva l'ordinamento delle modalità, ma non ne quantifica la distanza.

- Qualitativo sconnesso : grafico a torta

x_i	n_i	α_i°
lic. classico	30	62,43°
lic. scientifico	39	81,16°
artistico	16	33,30°
ragioneria	46	95,72°
ist. prof.	42	87,40°
	173	360°

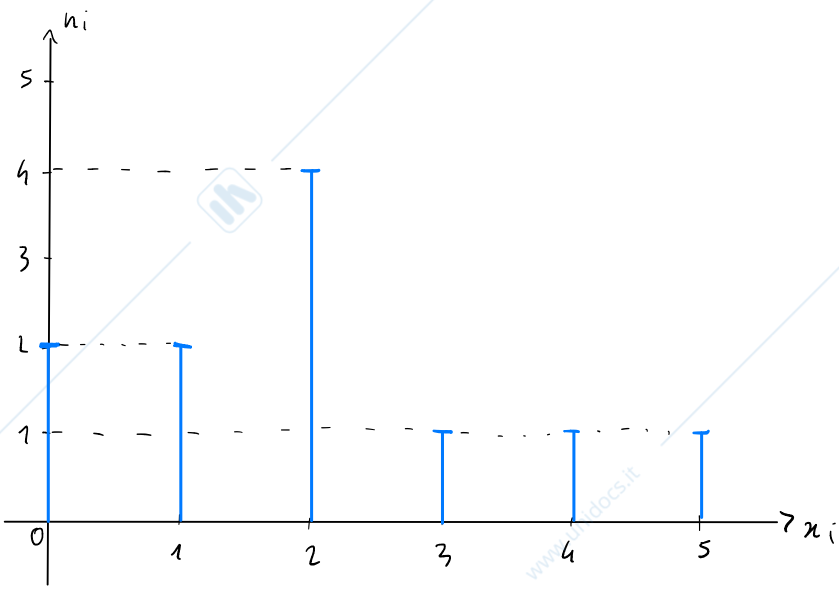


$\alpha_i^\circ = \quad n_i : n = \alpha_i^\circ : 360^\circ \qquad \alpha_i^\circ = \dfrac{n_i \cdot 360^\circ}{n}$

- Quantitativo discreto : grafico a bastoncini

x_i	n_i
0	2
1	2
2	4
3	1
4	1
5	1
	11

ordinabile => piano cartesiano



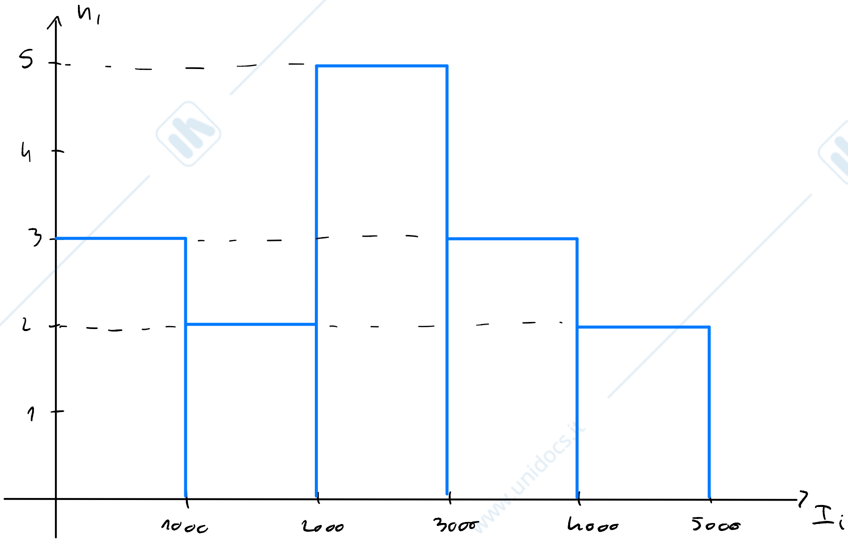
- Quantitativo continuo : istogramma (o "a canne d'organo")

NB classi di pari ampiezza : plottiamo le n_i
classi di ampiezza diversa : plottiamo le d_i

- Pari ampiezza

I_i	n_i
0 ⊣ 1000	3
1000 ⊣ 2000	2
2000 ⊣ 3000	5
3000 ⊣ 4000	3
4000 ⊣ 5000	2
	15

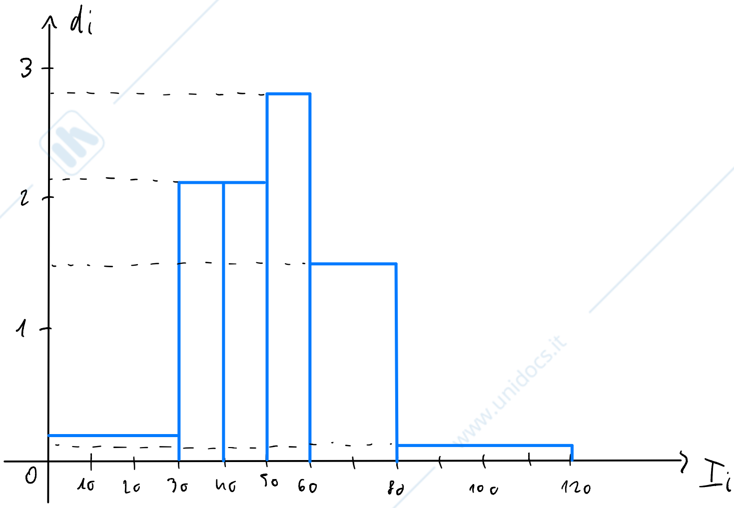
$a_i = 10$ costante



- Diversa ampiezza

I_i	n_i	a_i	d_i
0 ⊣ 30	6	30	0,2
30 ⊣ 40	21	10	2,1
40 ⊣ 50	21	10	2,1
50 ⊣ 60	27	10	2,7
60 ⊣ 80	29	20	1,45
80 ⊣ 120	2	40	0,05
	106		

Ampiezza diverse



Notare che se qui avessimo usato le n_i, la classe 60 ⊣ 80 sarebbe stata la più "alta".

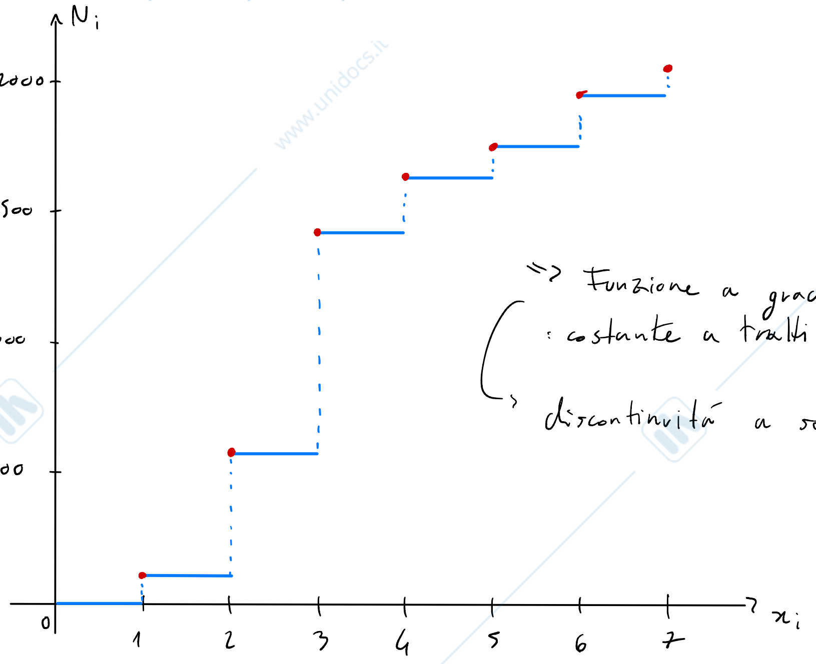
Grafico di frequenze cumulate

: si fa solo per caratteri quantitativi! (Per quelli qual. ordinabili ha senso calcolarle, ma non plottarle).

NB Anche qui N_i o F_i cambia solo la scala.

- Quantitativo discreto

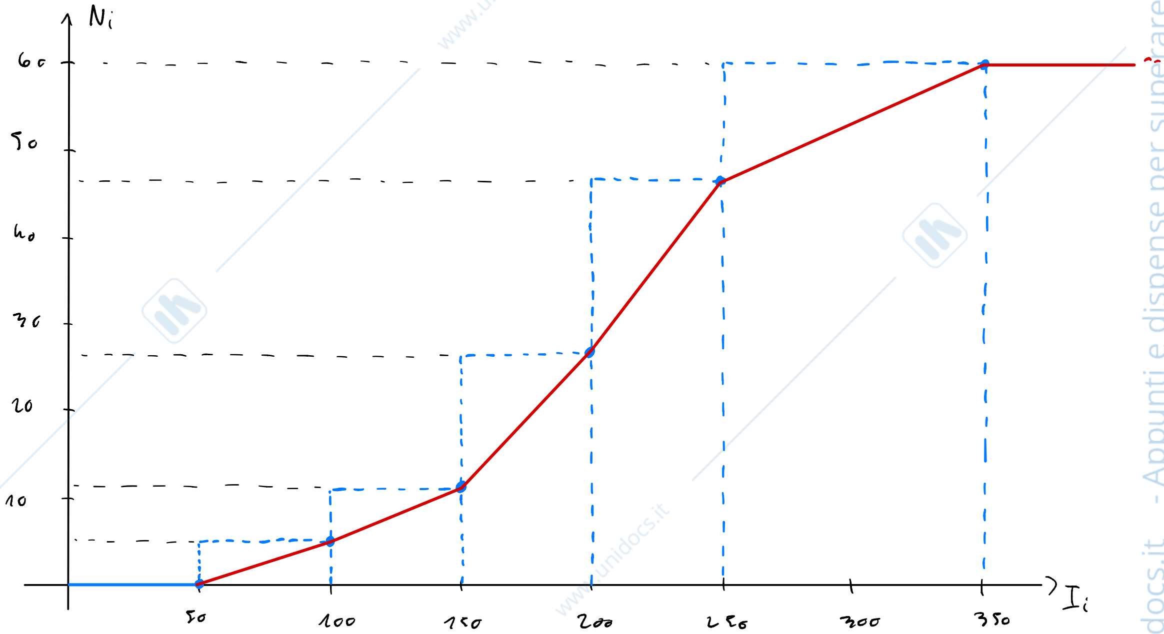
x_i	n_i	N_i
1	157	157
2	410	567
3	890	1457
4	215	1672
5	120	1792
6	150	1942
7	95	2037
	2037	



=> Funzione a gradini:
- costante a tratti
- discontinuità a salto

- Quantitativo continuo (siccome cumulo me ne frego delle ampiezze diverse)

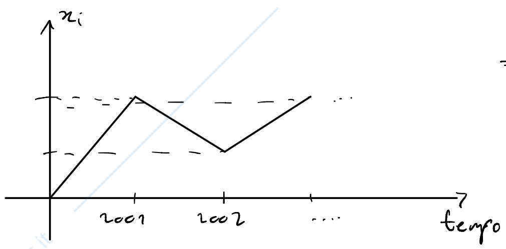
I_i	n_i	N_i
50 ⊣ 100	5	5
100 ⊣ 150	6	11
150 ⊣ 200	15	26
200 ⊣ 250	20	46
250 ⊣ 350	14	60
	60	



Siccome non so come sono distribuite le unità all'interno della classe, ipotizzo siano equidistribuite => unisco con la spezzata (che potrei interpolare per ottenere una curva polinomiale).

Altri grafici (cenni)

- Serie storica



=> studia l'andamento del fenomeno nel tempo

- Ideogrammi: carini da vedere ma difficili da leggere
- Cartogrammi: usano una cartina geografica, con diversi colori definiti da una legenda

Le Medie

: sintesi dei dati con la minore perdita d'informazione possibile

: una media é un indice sintetico "rappresentativo" delle modalitá assunte dalla variabile/mutabile

Def. sia X una variabile statistica tale che:

$$X = \{x_i ; n_i\} \text{ con } i = 1, \dots, r .$$

La MEDIA di X é una qualsiasi funzione reale α tale che:

$$\alpha = \alpha(X) = \alpha\{x_1, \dots, x_r ; n_1, \dots, n_r\}$$

che soddisfa le seguenti condizioni/proprietá:

1. Condizione di Cauchy (o "di internalitá")

$$x_1 \leq \alpha(X) \leq x_r \quad \text{dove} \quad x_1 = x_{min} \; , \; x_r = x_{max}$$

2. Proprietá moltiplicativa (c = costante)

$$\alpha(cx_1, cx_2, \dots, cx_r ; n_1, \dots, n_r) = \alpha(cx_i ; n_i) = \alpha(cX) = c \cdot \alpha(X)$$

3. Proprietá di monotonicitá

$$\text{se } x_i \geq y_i \;\; \forall i \;\; \text{allora} \;\; \alpha(X) \geq \alpha(Y)$$

[Una media é funzione monotona non decrescente dei valori del carattere.]

Una media "in senso stretto" soddisfa tutte e 3 le proprietá. Se mancano la 2 o la 3 o entrambe si parla di media "in senso lato", mentre la 1 deve sempre essere rispettata.

Moda

: indice di posizione NON analitico (non ha una formulazione matematica rigorosa => si puó usare quasi sempre)

: é comunque considerata una media, ma non analitica.

Def. Sia X una variabile o mutabile statistica qualsiasi

allora la MODA di X é: $Mo(X) = \tilde{x} = \{x_j : n_j = \max n_i\}$

Quindi la moda coincide con la (una) modalitá a cui é associata la massima frequenza (assoluta/relativa/densitá).

Nel caso di variabili quantitative continue, la classe I_i a cui é associata la massima densitá di frequenza sará la CLASSE MODALE:

$$\text{Classe modale} \quad I_j : d_j = \max d_i$$

La MODA allora coinciderá con il valore centrale della classe modale I_j:

$$Mo(x) = \frac{x_{max} - x_{min}}{2}$$ dove x_{max} e x_{min} sono gli estremi della classe modale.

La Moda puó quindi essere calcolata con qualsiasi tipo di carattere.

Lezione 6 -

Proprietá della Moda

① Cauchy : la Moda la rispetta sempre : $x_{min} \leq \tilde{x} \leq x_{max}$

② Moltiplicativa : la Moda la rispetta sempre (anche se ovviamente non si applica alle mutabili)

③ Monotonicitá : non sempre rispettata, in quanto la monotonicitá riguarda il valore delle modalitá, mentre la Moda le loro frequenze associate.

es.

$$X = \{1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4\}$$
$$Y = \{1, 1, 2, 2, 2, 2, 3, 3, 3, 4, 4, 4\}$$

$\Rightarrow y_{10} > x_{10} \;,\; y_{11} > x_{11}$

x_i	n_i
1	2
2	4
3	5
4	1
	12

y_i	n_i
1	2
2	4
3	3
4	3
	12

$\Rightarrow Mo(X) = 3 \quad > \quad Mo(Y) = 2$

$\Rightarrow$ Nonostante $y_i > x_i$, $Mo(X) > Mo(Y)$ $\Rightarrow$ Monotonicitá violata

Esempio

x_i Punti di vendita provinciale	n_i
Varese	141
Como	247 *
Sondrio	78
Milano	249 *
Bergamo	250 *
Brescia	206
Pavia	190
Cremona	115
Mantova	70
Totale	1546

$Mo(X) =$ Bergamo

NB caso considerabile come POLIMODALE : più di una Moda.

⇒ formalmente la $Mo(x)$ é Bergamo, ma Como e Milano presentano una differenza di fatto irrilevante considerata la mutabile in questione (qualitativa sconnessa).

NB si nota il carattere incerto che assume il concetto di Moda come "modalitá tipica".

Esempio

X = paghetta (⇒ quantitativo continuo)

I_i	n_i	a_i	d_i
0 ⊣ 8	140	8	17,5
8 ⊣ 12	136	4	34
12 ⊣ 18	180	6	30
18 ⊣ 24	150	6	25
24 ⊣ 28	90	4	22,5
	696		

Classe modale = 8 ⊣ 12

$$Mo(X) = \frac{12+8}{2} = 10$$

Mediana

: indice di posizione NON analitico

: calcolabile solo con caratteri ordinabili (⇒ NO qualitativi sconnessi)

: é una media in senso stretto

Def. La Mediana di X corrisponde al valore assunto da X che BIPARTISCE la distribuzione:

$$Me(X) = \hat{x} = \left\{ \hat{x} : \left[\sum_{x_i \leq \hat{x}} f_i \geq 0{,}50 \right] \wedge \left[\sum_{x_i \geq \hat{x}} f_i \geq 0{,}50 \right] \right\}$$

Quindi: la $Me(X)$ é quel valore superato e non superato da almeno il 50% delle modalitá.

Calcolo: - n pari: $\frac{n+1}{2}$ = posizione occupata da $Me(x)$

[= $Me(x)$ in caso di frequenze unitarie]

- n pari: $\frac{n}{2} = a$, $\frac{n}{2}+1 = b$

$\frac{a+b}{2}$ = posizione occupata da $Me(x)$ [oppure]

es.

x_i	n_i	f_i	F_i	R_i
1	4	0,2	0,2	1
2	5	0,25	0,45	0,8
3	3	0,15	0,60	0,55
4	6	0,3	0,90	0,4
5	2	0,1	1	0,1
	20	1		0

R_i = Retrocumulate : si parte da 1 e si sottraggono le f_i

$x_i = 3$ "contiene" il 50% delle modalità sia con le F_i che con le R_i

$\Rightarrow x_i = 3$ supera e non supera il 50% delle modalità

$\Rightarrow Me(X) = 3$

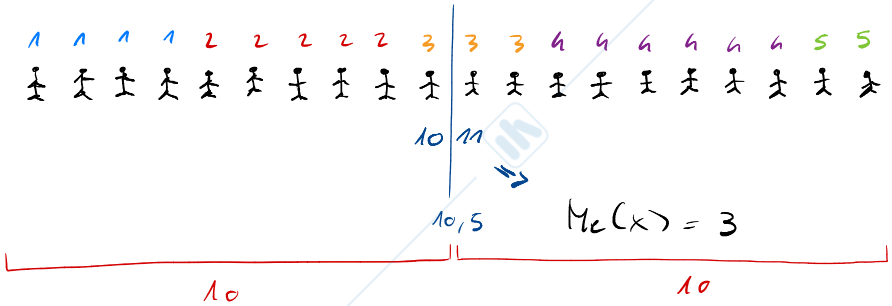
Metodo alternativo (per n pari) = semisomma

$\Rightarrow n = 20$ = pari quindi non c'è un unico valore centrale che bipartisce la distribuzione

$\Rightarrow$ semisomma dei 2 valori centrali che bipartiscono

$\frac{n}{2} = \frac{20}{2} = 10 = a$ $\qquad \frac{n}{2} + 1 = 11 = b$ $\qquad \frac{a+b}{2} = \frac{10+11}{2} = 10,5$

$\Rightarrow Me(X)$ sarà la modalità che occupa la posizione 10,5



$\Rightarrow Me(X) = 3$

Esempio

x_i	n_i	f_i	F_i	R_i
INS.	6	0,207	0,207	1
SUFF.	10	0,345	0,552	0,793
BUONO	8	0,276	0,828	0,448
OTTIMO	5	0,172	1	0,172
	29	1		0

$Me(X)$ = SUFF

$\Rightarrow$ comprende in entrambi i casi almeno il 50% della distribuzione

Attenzione : nel caso di caratteri quantitativi continui, i valori mediani sono potenzialmente infiniti, ma noi vogliamo un numero.

Si deriva allora geometricamente una formula.

Formula: $Me(X) = L_{inf} + \dfrac{0,5 - F_{inf}}{f_i} \cdot a_i$

dove: L_{inf} = limite inferiore della classe mediana

F_{inf} = F_i della classe precedente a quella mediana

f_i = f_i della classe mediana

a_i = ampiezza della classe mediana

Esempio

I_i	n_i	a_i	f_i	F_i	R_i
0 ⊣ 10	20	10	0,357	0,357	1
10 ⊣ 20	14	10	0,25	0,607	0,6427
20 ⊣ 30	9	10	0,1607	0,7678	0,3927
30 ⊣ 40	6	10	0,107	0,875	0,232
40 ⊣ 50	7	10	0,125	1	0,125
	56		1		0

Classe mediana = 10 ⊣ 20

$$Me(X) = 10 + \frac{0,5 - 0,357}{0,25} \cdot 10 = 15,72$$

Quantili e Percentili

generalizzazione della Mediana

Def. Dato p : $0 < p < 1$, il P-PERCENTILE (o p-quantile) x_p è il valore che delimita il primo p% delle osservazioni (ordinate in senso crescente) e le separa dal restante $(1-p)$%:

$$x_p = \left\{ x_p : \left[\sum_{x_i \le x_p} f_i \ge p \right] \wedge \left[\sum_{x_i \ge x_p} f_i \ge (1-p) \right] \right\}$$

es. $p = 0,5$ $\quad x_{0,5}$ = Mediana : 5° percentile = 2° quartile

$p = 0,25$ $\quad x_{0,25}$ = 1° quartile = modalità che supera almeno il 25% delle modalità ed è superata da almeno il 75% delle stesse.

$p = 0,1$ $\quad x_{0,1}$ = 10° percentile = 1° decile

$p = 0,75$ $\quad x_{0,75}$ = 3° quartile

⋮

Medie algebriche o "potenziate"

: indici di posizione analitici

: solo per variabili quantitative continue

Formula generica : $\mu_{(k)} = \left(\frac{1}{n} \sum_{i=1}^{m} n_i x_i^k \right)^{1/k} = \left(\sum_{i=1}^{n} f_i x_i^k \right)^{1/k}$

NB : $x_i > 0$

Medie potenziate ricorrenti :
- $k = 1$ Media aritmetica
- $k = 2$ Media quadratica
- $k = -1$ Media armonica
- $\lim_{k \to 0} \mu_k \to \mu_0$ Media geometrica

Media aritmetica ($k = 1$)

$$\mu_{(1)} = \frac{\sum_{i=1}^{m} x_i n_i}{n} = \sum_{i=1}^{m} f_i x_i$$

NB : $x_i > 0$ é inessenziale

Media armonica ($k = -1$)

$$\mu_{(-1)} = \left(\sum_{i=1}^{m} f_i x_i^{-1} \right)^{-1} = \frac{1}{\sum_{i=1}^{m} f_i \cdot \frac{1}{x_i}}$$

NB : $x_i \neq 0$ NB : $x_i > 0$

rischio di compensazione delle x_i con conseguente valore nullo della sommatoria

=> non si puó calcolare la $\mu_{(-1)}$ per X che assumono valori negativi o nulli (es. temperatura)

Media quadratica ($k = 2$)

$$\mu_{(2)} = \left(\sum_{i=1}^{m} f_i x_i^2 \right)^{1/2} = \sqrt{\sum_{i=1}^{m} f_i x_i^2}$$

NB : $x_i > 0$ non necessaria

NB : Date x_i uguali : $\mu_{(-1)} \leq \mu_{(1)} \leq \mu_{(2)}$

Lezione 7 - 18/10/21

le medie potenziate sono medie in senso stretto : rispettano le 3 proprietà

① Cauchy : $x_{min} \le \mu_{(k)} \le x_{max}$ <= conseguenza dell'attribuzione alle x_i di un sistema di pesi (p_i) che somma a 1

② Moltiplicatività : $\mu_{(k)}(c \cdot x) = \left[\sum_{i=1}^{m} (c \cdot x_i)^k p_i\right]^{1/k} = c \cdot \left(\sum_{i=1}^{m} x_i^k p_i\right)^{1/k}$
$= c \cdot \mu_{(k)}(x)$

③ Monotonicità rispetto agli x_i

per $k > 0$: $y_i \ge x_i \Rightarrow y_i^k \ge x_i^k$ data la crescenza monotonica di $f(x) = a^x$

per $k < 0$: $\mu_{(k)}(x) = \dfrac{1}{\left(\sum_{i=1}^{m} p_i \cdot \frac{1}{x_i^{|k|}}\right)^{1/|k|}} = \dfrac{1}{\mu_{(|k|)}(1/x)}$

allora $y_i \ge x_i \Rightarrow \dfrac{1}{y_i} \le \dfrac{1}{x_i} \Rightarrow \mu_{(|k|)}(1/y) \le \mu_{(|k|)}(1/x)$

$\Rightarrow \mu_{(|k|)}(y) \ge \mu_{(|k|)}(x)$

Teo. fondamentale sulle medie potenziate

Data $\mu_{(k)} = \left(\sum_{i=1}^{m} x_i^k p_i\right)^{1/k}$ $k \in [\pm 1, \pm 2, \pm 3, \dots]$ e x_i positivi ordinati e distinti

① Monotona in senso stretto (rispetto a k)

$\mu_{(s)} \ge \mu_{(r)}$ se $s \ge r$

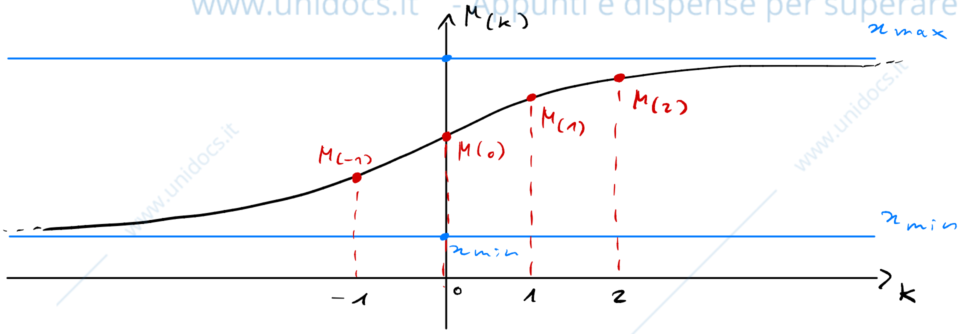
② $\lim_{k \to +\infty} \mu_{(k)} = x_{max}$; $\lim_{k \to -\infty} \mu_{(k)} = x_{min}$

=> funzione Media potenziata ha due asintoti orizzontali in corrispondenza di $y = x_{min}$ e $y = x_{max}$

③ Ogni media potenziata di ordine k (con k qualsiasi) è media in senso stretto

④ Media geometrica

$\lim_{k \to 0} \mu_{(k)} = \sqrt{x_1^{m_1} \cdot x_2^{m_2} \cdot \dots \cdot x_m^{m_m}} = \sqrt[m]{\prod_{i=1}^{m} x_i^{m_i}}$ = Media geometrica



$\Rightarrow$ é chiaro come a un certo punto gli ordini perdano di interesse

DIM.

① Troppo avanzata

② valore più grande osservato

$$\lim_{k\to+\infty} M_{(k)} = \lim_{k\to+\infty}\left(\frac{\sum_{i=1}^{r} x_i^k \cdot m_i}{n}\right)^{1/k} = \lim_{k\to+\infty}\left(\frac{\overset{x_{max}}{x_r^k} \cdot m_r}{n} + \frac{\sum_{i=1}^{r-1} x_i^k \cdot m_i}{n}\right)^{1/k} =$$

$$= \lim_{k\to+\infty}\left[x_r^k\left(\frac{x_r^k}{x_r^k}\cdot\frac{m_r}{n} + \sum_{i=1}^{r-1}\frac{x_i^k}{x_r^k}\cdot\frac{m_i}{n}\right)\right]^{1/k} =$$

$$= \lim_{k\to+\infty} x_r\left[\frac{m_r}{n} + \sum_{i=1}^{r-1}\left(\frac{x_i^k}{x_r^k}\cdot\frac{m_i}{n}\right)\right]^{1/k} = x_r \cdot \lim_{k\to+\infty}\left[\overset{ii}{\frac{m_r}{n}} + \overset{i}{\sum_{i=1}^{r-1}\left(\frac{x_i^k}{x_r^k}\cdot\frac{m_i}{n}\right)}\right]^{1/k}$$

i) $x_i < x_r \ \forall i<r \Rightarrow x_i^k < x_r^k \ \forall i<r \Rightarrow \frac{x_i^k}{x_r^k} < 1$

$$\Rightarrow \lim_{k\to+\infty}\left(\frac{x_i}{x_r}\right)^k = 0 \Rightarrow (0)^{1/k} = 0$$

ii) $x_r \cdot \lim_{k\to+\infty}\left(\frac{m_r}{n}\right)^{1/k \to 0} = 1$

$m_r > 0$; $m_r < n \Rightarrow 0 < \frac{m_r}{n} < 1$

$$\Rightarrow x_r \cdot (1 + 0) = x_r$$

$\Rightarrow \lim_{k\to+\infty} M_{(k)}$ tende al valore più grande osservato

NB: per $K\to-\infty$ la DIM é identica, semplicemente si mette in evidenza $x_1 = x_{min}$ e si lavora con il $-\infty$

③ Giá dimostrata

④ $$\lim_{k\to 0} M_{(k)} = \lim_{k\to 0}\left(\sum_{i=1}^{m} \overset{\to 1}{x_i^k}\cdot f_i\right)^{1/k \to \infty} = 1^{\infty} \quad \text{F.D.I.}$$

$$\lim_{k\to 0} M_{(k)} = M_g(x) \qquad \ln M_g(x) = \ln \lim_{k\to 0} M_{(k)} = \lim_{k\to 0} \ln M_{(k)} =$$

$$= \lim_{k\to 0} \ln\left(\sum_{i=1}^{r} x_i^k \cdot f_i\right)^{1/k} = \lim_{k\to 0} \frac{1}{k}\ln\left(\sum_{i=1}^{r} x_i^k \cdot f_i\right) = \frac{0}{0} \quad \text{FDI}$$

NB: $\ln a^x = x \ln a$

De L'Hospital : $\dfrac{dN/dk}{dD/dk}$

i) D : k $\quad \dfrac{dD}{dk} = \dfrac{d}{dk} k = 1 \quad \left[\Leftarrow \dfrac{dk}{dk} = 1 \right]$

ii) N : $\ln \left(\sum_{i=1}^{r} x_i^k \cdot f_i \right) \qquad \dfrac{dN}{dk} = \dfrac{d}{dk} \ln \sum_{i=1}^{r} \left(x_i^k \cdot f_i \right)$ NB : $\dfrac{d}{dx} \ln f(x) = \dfrac{f'(x)}{f(x)}$

$= \dfrac{\sum_{i=1}^{r} \frac{d}{dk} x_i^k \cdot f_i}{\sum_{i=1}^{r} x_i^k \cdot f_i} =$ NB : $\dfrac{d}{dx} a^x = \ln a \cdot a^x \quad = \dfrac{\sum_{i=1}^{r} f_i \left[(\ln x_i) \cdot x_i^k \right]}{\sum_{i=1}^{r} x_i^k \cdot f_i}$

$\Rightarrow \ln M_g(x) = \lim_{k \to 0} \dfrac{\sum_{i=1}^{r} f_i \left[(\ln x_i) \cdot x_i^k \right]}{\sum_{i=1}^{r} x_i^k \cdot f_i} = \dfrac{\lim_{k \to 0} \sum_{i=1}^{r} f_i \left[(\ln x_i) \cdot x_i^k \right]}{\lim_{k \to 0} \sum_{i=1}^{r} f_i \cdot x_i^k} =$

$= \dfrac{\sum_{i=1}^{r} f_i \cdot \ln(x_i)}{\sum_{i=1}^{r} f_i} = \sum_{i=1}^{r} f_i \cdot \ln(x_i) = \sum_{i=1}^{r} \ln (x_i)^{f_i}$ NB : $\ln \prod_{i=1}^{r} \ldots = \sum_{i=1}^{r} \ln \ldots$

$= \ln \prod_{i=1}^{r} (x_i)^{f_i} = \ln M_g(x)$: faccio l'exp per ottenere $M_g(x)$

$\Rightarrow e^{\ln M_g(x)} = e^{\ln \prod_{i=1}^{r} (x_i)^{f_i}} \Rightarrow M_g(x) = \prod_{i=1}^{r} (x_i)^{f_i} = \prod_{i=1}^{r} (x_i)^{\frac{n_i}{n}} = \sqrt[n]{\prod_{i=1}^{r} (x_i)^{n_i}}$

Media aritmetica

- Indice di posizione : esprime su quale punto dell'asse reale è concentrata la distribuzione.

- Si indica con μ $(= \mu_{(1)})$

Proprietà esclusive (soddisfa le 3 delle medie)

① $\sum_{i=1}^{r} (x_i - \mu) \cdot n_i = 0$

② $\sum_{i=1}^{r} (x_i - \mu)^2 n_i = \min$

Proprietà dell'operatore Media

- $M(x)$ in ITA, $E(x)$ in ENG

Prendiamo $a \in \mathbb{R}$, $b \in \mathbb{R}$

① $M(a) = a$

② $M(a\,x) = a\,M(x)$

$$M(a\,x) = \frac{\sum_{i=1}^{r} a x_i n_i}{n} = a \frac{\sum_{i=1}^{r} x_i \cdot n_i}{n} = a\,M(x)$$

③ $M(x+y) = M(x) + M(y)$

$$M(x+y) = \frac{\sum_{i=1}^{r} (x_i + y_i) n_i}{n} = \frac{\sum_{i=1}^{r} x_i n_i}{n} + \frac{\sum_{i=1}^{r} y_i n_i}{n}$$

④ $M(a + bx) = a + b\,M(x)$

$$\frac{\sum_{i=1}^{r} (a + b x_i) n_i}{n} = \frac{\sum a n_i}{n} + \frac{\sum b x_i n_i}{n} = a + b \frac{\sum x_i n_i}{n}$$

due proprietà esclusive della Media aritmetica

DIM (1)

$$\sum_{i=1}^{r} (x_i - \mu) n_i = 0 \quad \text{oppure} \quad \frac{\sum_{i=1}^{r} (x_i - \mu) n_i}{n} = \sum_{i=1}^{r} (x_i - \mu) f_i =$$

$$= \sum_{i=1}^{r} x_i f_i - \sum_{i=1}^{r} \mu f_i = \mu - (\mu \cdot 1) = 0$$

DIM. (2)

$$\sum_{i=1}^{r} (x_i - \mu)^2 n_i = \min$$

con derivate (ne calcolo la derivata e la pongo $= 0$)

metodo analitico

metodo analitico: $\min_{\theta} \sum_{i=1}^{r} (x_i - \theta)^2 = ??$ Quale θ rende minima questa quantità?

$\sum_{i=1}^{r} (x_i - \theta)^2$ NB: tolgo f_i, tanto sono solo un sistema di pesi $= \sum_{i=1}^{r} \left[(x_i - \theta + \mu - \mu)^2 \right] =$

$$= \sum_{i=1}^{r} \left[(x_i - \mu) - (\theta - \mu) \right]^2 = \sum_{i=1}^{r} \left[(x_i - \mu)^2 + (\theta - \mu)^2 - 2 (x_i - \mu)(\theta - \mu) \right]$$

$$= \sum_{i=1}^{r} (x_i - \mu)^2 + \sum_{i=1}^{r} (\theta - \mu)^2 - 2 \sum_{i=1}^{r} (x_i - \mu)(\theta - \mu) =$$

$$\sum_{i=1}^{r}(x_i-\mu)^2 + n(\theta-\mu)^2 - 2(\theta-\mu)\underbrace{\sum_{i=1}^{r}(x_i-\mu)}_{=0} =$$

$$= \sum_{i=1}^{r}(x_i-\mu)^2 + n(\theta-\mu)^2 \overset{?}{=} \min$$

Con $\theta = \mu \Rightarrow n(\theta-\mu)^2 = 0$ CVD

Ancora sulle medie potenziate:

Abbiamo detto che sono infinite, ma ce ne interessano solo 4 o 5.
Ma come si sceglie quale usare?

Un paio di criteri:

① Metodo del Chisini o Media obiettivo

(non sempre restituisce una media in senso stretto)

Data X, Impongo un vincolo globale (= funzione obiettivo)

$$\varphi(X) = \varphi(x_1, x_2, \dots, x_r; n_1, \dots, n_r)$$

CHISINI: risolvere l'equazione

$$\varphi(x_1, \dots, x_r; n_1, \dots, n_r) = \varphi(\alpha, \dots, \alpha; n_1, \dots n_r)$$

sostituisco la costante α agli x_i

Se φ è monotona continua avrà una soluzione unica.

Tipiche funzioni obiettivo:

- $\varphi(x) = \sum_{i=1}^{r} x_i n_i$ Funzione additiva (es. costo totale, ...)

$$\Rightarrow \sum_{i=1}^{r} x_i n_i = \sum_{i=1}^{r} \alpha n_i \Rightarrow \alpha = \frac{\sum_{i=1}^{r} x_i n_i}{n}$$ Totale invariato

- $\varphi(x) = \prod_{i=1}^{r} x_i^{n_i}$ Funzione moltiplicativa

$$\Rightarrow \prod_{i=1}^{r} x_i^{n_i} = \prod_{i=1}^{r} \alpha^{n_i} \Rightarrow \prod_{i=1}^{r} x_i^{n_i} = \alpha^{n} \quad (n = \Sigma n_i) \Rightarrow \alpha = \sqrt[n]{\prod_{i=1}^{r} x_i^{n_i}} = M_g(X)$$

Prodotto globale invariato