8. GLI INDICI DI POSIZIONE (o "di centralità")

A) LE MEDIE "FERME"

(una media si dice "**ferma**" se **il suo valore varia senz'altro, qualora uno solo dei termini in gioco cambi**).

- HO COMINCIATO AD ALLENARMI PER LA MARATONA DI NEW YORK!

  Lunedì ho corso per 4 km, martedì per 6,5 km, mercoledì per 8 km, giovedì per 2,5 km, venerdì per 5 km, sabato per 7,5 km, domenica per 8,5 km.
  Quanti km ho percorso in media al giorno?

  Una "**media**" **fra più numeri**, che esprimono quantità della stessa specie, **è un numero avente la proprietà di essere compreso ("medio" = "che sta in mezzo") tra il minore e il maggiore dei numeri dati**.

  La risposta alla nostra domanda NON sarà però, evidentemente,
  uno qualsiasi fra i valori compresi tra 2,5 e 8,5
  e nemmeno *il numero esattamente intermedio* fra 2,5 e 8,5 (che sarebbe 5,5)!

  Ragioniamo. Noi vogliamo trovare quel numero x tale che,
  se in ognuno dei 7 giorni della settimana io avessi corso ogni volta per esattamente x km,
  la distanza complessiva percorsa in tutta la settimana sarebbe stata la medesima! Allora

  $$7x = 4+6,5+8+2,5+5+7,5+8,5$$

  da cui

  $$x = media\ dei\ km = n°\ di\ km\ \ percorsi\ mediamente\ in\ 1\ giorno = \frac{4+6,5+8+2,5+5+7,5+8,5}{7} = \frac{42}{7} = 6$$

  In effetti, se ogni giorno della settimana il mio percorso fosse stato di esattamente 6 km,
  complessivamente nella settimana mi sarei allenato per un totale di $6+6+6+6+6+6+6 = 42$ km,
  esattamente come è $4+6,5+8+2,5+5+7,5+8,5 = 42$.

- E' ragionevole supporre che la prestazione complessiva di uno studente, che ha preso diversi voti

  $$v_1, v_2, ..., v_n,$$

  possa essere bene rappresentata dal particolare voto v che, se fosse stato preso, *sempre quello*, tutte le n volte, avrebbe dato luogo alla *stessa somma di voti*. Dunque

  $$nv = v_1 + v_2 + ... + v_n$$

  $$v = media\ dei\ voti = \frac{v_1 + v_2 + ... + v_n}{n}$$

  Ad es., se quello studente ha preso i 5 voti $6\ \ 7\ \ 7\ \ 8\ \ 6$, la sua media è stata $\frac{6+7+7+8+6}{5} = \frac{34}{5} = 6,8$.

  E in effetti, se quello studente avesse invece preso come voti successivi $6,8\quad 6,8\quad 6,8\quad 6,8\quad 6,8$
  la somma dei suoi voti, ossia la sua "prestazione totale", sarebbe stata $6,8+6,8+6,8+6,8+6,8 = 34$
  esattamente uguale alla prestazione effettiva totale $6+7+7+8+6 = 34$

  Si dice **MEDIA ARITMETICA** M fra n numeri $x_1, x_2, ..., x_n$, il numero

  $$\boxed{\boldsymbol{M = media\ aritmetica = \frac{x_1 + x_2 + ... + x_n}{n}}}.$$

  La media aritmetica fra più valori, è uguale alla loro somma, divisa per il numero dei valori stessi; ed è quel nuovo valore il quale, se sostituito al posto di ciascuno dei singoli valori in gioco, ne lascerebbe invariata la somma.

ESERCIZI

1) Verifica che se ho mangiato il minestrone 3 volte in novembre, 8 in dicembre, 8 in gennaio, 5 in febbraio, e 9 volte in totale gli altri mesi dell'anno, in media è come se l'avessi mangiato 2,75 volte al mese.

2) Calcola la media del numero di scarpe, fra i compagni di classe: a) maschi; b) femmine; c) tutti.
   In generale la media c) *non* coincide con la media delle due medie a) e b) … a meno che …

3) Con riferimento alle classi I A, I B di cui a pagina 354, con un foglio elettronico rappresenta la "serie storica" delle medie aritmetiche dei punteggi (mettendo, ogni anno, assieme le due classi a formare un unico gruppo).
   Le "lagnanze" di cui si parla nella stessa pagina sono giustificate, a giudicare da questa successione di medie?

La "media" di cui ci siamo occupati fin qui è stata la media "aritmetica"
(anche se sovente, per brevità, l'aggettivo viene lasciato sottinteso);
in effetti, **ci sono altri tipi di "medie"**, oltre a questa, e ora andremo brevemente a illustrarli.

- Se il COSTO DI UNA MATERIA PRIMA è aumentato:
  - del 5% nel 2001 (s'intende: da inizio a fine anno);
  - del 6% nel 2002;
  - dell'8% nel 2003;
  - dell'8% ancora nel 2004;
  - e del 4% nel 2005,

  a) di quanto è aumentato *complessivamente* nel quinquennio 2001-2005?

  b) E di quanto è aumentato *mediamente* ogni anno, in questo quinquennio?

Ragioniamo.

a) Se il prezzo all'inizio del 2001 era 100,

- alla fine del 2001 è diventato $100 \cdot \left(1+\frac{5}{100}\right) = 100 \boxed{\cdot 1{,}05} = 105$
- alla fine del 2002 è diventato $105 \cdot \left(1+\frac{6}{100}\right) = 105 \boxed{\cdot 1{,}06} = 111{,}3$
- alla fine del 2003 è diventato $111{,}3 \cdot \left(1+\frac{8}{100}\right) = 111{,}3 \boxed{\cdot 1{,}08} = 120{,}2$ (*circa*)
- alla fine del 2004 è diventato *circa* $120{,}2 \cdot \left(1+\frac{8}{100}\right) = 120{,}2 \boxed{\cdot 1{,}08} =$ *circa* $129{,}8$
- alla fine del 2005 è diventato *circa* $129{,}8 \cdot \left(1+\frac{4}{100}\right) = 129{,}8 \boxed{\cdot 1{,}04} =$ *circa* 135;

è perciò aumentato, questo prezzo, da inizio 2001 a fine 2005, complessivamente intorno al 35%.

b) E *mediamente*, quanto è aumentato?
Noi *cerchiamo* in questo momento *una percentuale annua x tale che,*
se *l'aumento fosse stato ogni anno esattamente dell'*x%, *si sarebbe raggiunto il medesimo prezzo finale.*
Quindi

$$\cancel{100} \cdot \left(1+\frac{x}{100}\right)\left(1+\frac{x}{100}\right)\left(1+\frac{x}{100}\right)\left(1+\frac{x}{100}\right)\left(1+\frac{x}{100}\right) = \cancel{100} \cdot \left(1+\frac{5}{100}\right)\left(1+\frac{6}{100}\right)\left(1+\frac{8}{100}\right)\left(1+\frac{8}{100}\right)\left(1+\frac{4}{100}\right)$$

$$\left(1+\frac{x}{100}\right)^5 = 1{,}05 \cdot 1{,}06 \cdot 1{,}08 \cdot 1{,}08 \cdot 1{,}04$$

$$1+\frac{x}{100} = \sqrt[5]{1{,}05 \cdot 1{,}06 \cdot 1{,}08 \cdot 1{,}08 \cdot 1{,}04}$$

$$1+p = \sqrt[5]{1{,}05 \cdot 1{,}06 \cdot 1{,}08 \cdot 1{,}08 \cdot 1{,}04} \approx \sqrt[5]{1{,}35} \approx 1{,}062 \quad \textit{da cui} \quad p = \sqrt[5]{1{,}05 \cdot 1{,}06 \cdot 1{,}08 \cdot 1{,}08 \cdot 1{,}04} - 1 \approx 0{,}062 = 6{,}2\%$$

Vuol dire che, se quel prezzo iniziale fosse aumentato ogni anno del $6{,}2\%$, si sarebbe raggiunto,
dopo un quinquennio, lo stesso prezzo finale che si è ottenuto con gli aumenti del 5%, 6%, 8%, 8%, 4%.
Vuoi provare a verificarlo col calcolo?

Presi dunque i valori $1{,}05 \;\; 1{,}06 \;\; 1{,}08 \;\; 1{,}08 \;\; 1{,}04$
che davano il numero per cui moltiplicare il prezzo all'inizio dell'anno, onde ottenere il prezzo alla fine,
la "media sul quinquennio" di questi moltiplicatori è la radice quinta del loro prodotto
(e non, come nel caso della media aritmetica, la quinta parte della loro somma)!

**Si definisce "MEDIA GEOMETRICA" fra più valori, quel nuovo valore il quale,
se sostituito al posto di ciascuno dei singoli valori in gioco, ne lascerebbe invariato il prodotto.**
Si dimostra facilmente che la media geometrica fra *n* valori $x_1, x_2, \ldots, x_n$ è data da

$$\boldsymbol{M_G = \textit{media geometrica} = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}}$$

Infatti, poiché si desidera che questa media, sostituita a ciascuno dei valori, non ne alteri il prodotto:

$$x_1 \cdot x_2 \cdot \ldots \cdot x_n = M_G \cdot M_G \cdot \ldots \cdot M_G$$
$$M_G{}^n = x_1 \cdot x_2 \cdot \ldots \cdot x_n$$
$$M_G = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$$

Come indicazione generale, possiamo dire che la media geometrica
si utilizza quando i dati sono tali che per essi l' "operazione regina"
è il *prodotto*, piuttosto che la somma.
Quindi, in un contesto di tassi di interesse bancari,
o di aumento o diminuzione (rara …) dei prezzi,
o di incremento o decremento del PIL, o di tasso di crescita di una popolazione,
dobbiamo aspettarci di incontrare medie geometriche piuttosto che aritmetiche.

Robert Kennedy [*Discorso sul PIL*] *(marzo 1968)*

- Facciamo UN ALTRO ESEMPIO DI NATURA DIVERSA, molto significativo.

$$v=\frac{s}{t}\quad s=vt\quad t=\frac{s}{v}$$

Se ho percorso in auto un totale di 100 km, la prima metà andando ai 100 km/h
e la metà successiva, dopo aver visto un brutto incidente, agli 80 km/h soltanto,
quale è stata la mia velocità media?

Rispondere che è stata la media aritmetica delle due velocità,
quindi $\frac{100+80}{2}=90$ km/h, sarebbe magari istintivo … ma clamorosamente SBAGLIATO.

Infatti è logico partire dal presupposto che per "velocità media", in questo contesto, si debba intendere *quella velocità la quale, se mantenuta costante per tutto il tragitto di 100 km, mi avrebbe permesso di coprirlo nel medesimo tempo.*

E *quanto tempo* ci ho messo a fare i miei 100 km,
andando per 50 km ai 100 all'ora e per 50 km agli 80 all'ora?

Vediamo.
La prima metà del percorso ha richiesto un tempo, in ore, uguale a

$\frac{s}{v}=\frac{50}{100}=\frac{1}{2}=0,5$ (mezz'ora, dunque),

mentre la seconda metà ha richiesto un numero di ore dato da

$\frac{s}{v}=\frac{50}{80}=0,625$ (0,625 ore, o anche: 37 minuti e mezzo).

Il tempo totale per coprire il tragitto di 100 km è stato perciò di ore $0,5+0,625=1,125\ (1h\ 7'\ 30'')$.
Ma se una distanza di 100 km venisse percorsa ad andatura costante in ore 1,125
vorrebbe dire che quella velocità costante è di

$$\text{km/h}\ \frac{100}{1,125}\approx \text{km/h}\ 88,9$$

Quindi in questo caso per calcolare la "velocità media"
NON si deve fare la "media aritmetica delle due velocità"!

Si deve invece procedere
1) direttamente col ragionamento e col calcolo, come abbiamo fatto noi;
2) oppure (lo si potrebbe dimostrare) calcolando la cosiddetta *media armonica* delle velocità.

$$\boldsymbol{M_A = \textit{media armonica} = \frac{1}{\dfrac{\dfrac{1}{x_1}+\dfrac{1}{x_2}+\ldots+\dfrac{1}{x_n}}{n}}}$$

Si definisce "MEDIA ARMONICA" fra n valori, quel nuovo valore il quale, se sostituito al posto di ciascuno dei singoli valori in gioco, ne lascerebbe invariata la somma dei reciproci. Essa coincide col reciproco della media aritmetica dei reciproci dei valori in gioco.

Si può far vedere (vuoi provarci?) che, se una data distanza d viene suddivisa
in n tratti tutti uguali fra loro (ogni tratto ha quindi lunghezza d/n),
e questi tratti vengono percorsi alle velocità $v_1, v_2, \ldots, v_n$ rispettivamente,
per cui il viaggio richiede un certo tempo totale t,
allora la "velocità media", intesa come la velocità costante
alla quale occorrerebbe muoversi per percorrere la stessa distanza d nello stesso tempo t,

- *non dipende dalla distanza* d
- *ed è data dalla media armonica delle velocità:*

$$\boxed{v=\frac{1}{\dfrac{\dfrac{1}{v_1}+\dfrac{1}{v_2}+\ldots+\dfrac{1}{v_n}}{n}}}$$

OSSERVAZIONE
Invece, se noi avessimo un tempo di viaggio fissato t suddiviso in n intervalli di ugual durata t/n,
e in questi n intervalli uguali di tempo si procedesse alle velocità costanti $v_1, v_2, \ldots, v_n$,
percorrendo una determinata distanza totale d,
la velocità costante alla quale procedere se si desidera, sempre nel tempo t, percorrere la stessa distanza d
non dipenderebbe da t e sarebbe data dalla *media aritmetica* delle velocità $v'=\frac{v_1+v_2+\ldots+v_n}{n}$ (dimostralo!)

- Ancora:

  Si definisce "MEDIA QUADRATICA" fra più valori, quel nuovo valore il quale,
  se sostituito al posto di ciascuno dei singoli valori in gioco,
  ne lascerebbe invariata la somma dei quadrati.

  Si dimostra che la media quadratica fra n valori $x_1, x_2, \dots, x_n$ è data da

$$\boxed{M_Q = \textit{media quadratica} = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}}}$$

Cercando di trarre le conclusioni da questo tentativo di GENERALIZZAZIONE DEL CONCETTO DI "MEDIA", potremo dire (traducendo in forma più semplice una definizione di Oscar Chisini, 1889-1967) che
se si hanno n valori $x_1, x_2, \dots, x_n$ di una grandezza, si può parlare di "media" ogniqualvolta si desidera determinare un valore x che, qualora venisse sostituito al posto di ciascuno dei valori dati, ne lascerebbe invariata, a seconda del tipo di "media":

- **la somma;**
- **o il prodotto;**
- **…**
- **oppure una qualunque determinata loro "funzione",**
  ossia grandezza che dipenda, secondo una legge ben definita, dalle grandezze date.

UNA MEDIA
a) … IN PARTE DISTRUGGE, E IN PARTE RIESCE A MANTENERE L'INFORMAZIONE;
b) … E' UN VALORE "TEORICO";
c) … E CI DA' SOLO QUELLO CHE DA LEI SAPPIAMO DI POTERCI ASPETTARE!

a) Una media, di qualsiasi tipo essa sia, cerca di sintetizzare in un singolo numero un'informazione relativa a una pluralità di dati (sovente, a *tantissimi* dati).
Evidentemente, essa *non* può pretendere di condensare in sé tutto il contenuto informativo insito nell'insieme *effettivo* dei dati; passando alla "media" tale contenuto in gran parte va perso …
e tuttavia qualcosa, peraltro di molto importante, rimane.

b) Una media è un valore "TEORICO", nel senso che ben raramente coincide con uno dei dati in questione (e, se anche ciò avviene, questo fatto non è comunque particolarmente interessante).

c) Una media "CI DA' SOLO QUELLO CHE DA LEI CI ASPETTIAMO", nel senso che, ad esempio,
- una media *aritmetica* ci dà quel valore che, se venisse sostituito a ciascuno dei dati, ne lascerebbe inalterata la somma;
- una media *geometrica* ci dà quel valore che, se venisse sostituito a ciascuno dei dati, ne lascerebbe inalterato il prodotto;
- una media *armonica* ci dà quel valore che, se venisse sostituito a ciascuno dei dati, ne lascerebbe inalterata la somma dei reciproci;
- eccetera.

LA PIENA COMPRENSIONE DEL SIGNIFICATO DI UNA "MEDIA",
cioè del tipo di informazione che essa ci dà,
È LEGATA ALLA CONSAPEVOLEZZA DI
"QUAL È LA QUANTITÀ CHE RESTEREBBE INALTERATA
SE AL POSTO DI CIASCUNO DEI DATI
SI SOSTITUISSE LA MEDIA DEI DATI STESSI"

Si dice che la media in esame "CONSERVA"
quella determinata quantità:
ad esempio, la media aritmetica "conserva la somma",
perché, se venisse sostituita al posto di ciascuno dei dati,
la somma di questi non muterebbe.

- La media con cui lo studente ha quasi sempre a che fare è la media *aritmetica* (quando si dice semplicemente "*media*", *senza aggettivi*, è alla *media aritmetica* che ci si riferisce).
- Per la precisione, nelle pagine precedenti, avremmo dovuto scrivere "media aritmetica *semplice*", "media geometrica *semplice*", "media armonica *semplice*", … per distinguere le medie introdotte dalle corrispondenti medie "*ponderate*".
  Alla media aritmetica "ponderata" faremo cenno fra breve.

In un'indagine statistica, o in un diagramma statistico,
i "dati" di cui fare la "media" sono le "modalità"
(è ovvio che ha senso farne la media soltanto se queste sono espresse numericamente);
ciascuna modalità viene contata tante volte quant'è la sua frequenza nella popolazione statistica in esame.

Ad esempio, nella rilevazione del numero di figli da 0 a 10 anni di un gruppo di 20 famiglie,
la distribuzione di frequenze potrebbe essere

0	1	2	3
7	8	4	1

E in questo caso avrebbe senso fare la media aritmetica del numero dei figli, che sarebbe

$$M = \frac{0+0+0+0+0+0+0+1+1+1+1+1+1+1+1+2+2+2+2+3}{20} = \frac{19}{20} = 0,95$$

Se i dati sono stati ripartiti in intervalli ossia, come si dice, in "**CLASSI DI FREQUENZA**"
(ad esempio in una rilevazione di altezze: $cm\ 150 \le h < cm\ 154$, $154 \le h < 158$, $158 \le h < 162$, $ecc.$),
nel calcolo di una media si prende, per ciascuna classe, il cosiddetto "**VALORE CENTRALE**" **della classe**,
ossia la **semisomma** (= la media) **delle estremità dell'intervallo**. Esempio:

$150 \le h < 154$	$154 \le h < 158$	$158 \le h < 162$	$162 \le h < 166$	$166 \le h < 170$
2	5	8	10	15
$170 \le h < 174$	$174 \le h < 178$	$178 \le h < 182$	$182 \le h < 186$	$186 \le h < 190$
9	7	5	2	1

$$M = \frac{152\cdot 2+156\cdot 5+160\cdot 8+164\cdot 10+168\cdot 15+172\cdot 9+176\cdot 7+180\cdot 5+184\cdot 2+188\cdot 1}{2+5+8+10+15+9+7+5+2+1} = \frac{10760}{64} = 168,125$$

Nell'ultimo esempio anziché sommare un certo numero di addendi uguali
abbiamo moltiplicato ciascun addendo per il numero di volte in cui questo andava considerato;
abbiamo cioè fatto quella che, come vedremo poco più avanti, si chiama una "media PONDERATA".

E' evidente che questo metodo del "valore centrale" non fornisce come risultato la media "esatta",
ma solo un valore approssimato della "vera" media.
La "vera" media, infatti, dovrebbe tenere conto di *tutti* i singoli valori osservati
(che per comodità sono invece stati riuniti in classi);
ciascun singolo valore dovrebbe essere moltiplicato per la sua brava frequenza,
questi prodotti sommati e infine questa somma divisa per il numero totale dei valori considerati.
L'approssimazione però in genere è molto buona …
Rinunciamo ad ulteriori approfondimenti, ma possiamo comunque fare un "esperimento" pratico.
Riprendiamo la tabella precedente ed entriamo nel dettaglio delle singole osservazioni:

$150 \le h < 154$		$154 \le h < 158$		$158 \le h < 162$		$162 \le h < 166$		$166 \le h < 170$	
150	0	154	0	158	1	162	2	166	4
151	0	155	1	159	2	163	2	167	4
152	1	156	2	160	3	164	3	168	3
153	1	157	2	161	2	165	3	169	4
	2		5		8		10		15

$170 \le h < 174$		$174 \le h < 178$		$178 \le h < 182$		$182 \le h < 186$		$186 \le h < 190$	
170	3	174	3	178	2	182	1	186	0
171	2	175	1	179	1	183	0	187	1
172	2	176	2	180	0	184	1	188	0
173	2	177	1	181	2	185	0	189	0
	9		7		5		2		1

Facendo, questa volta, la "vera" media si ottiene un valore vicino a $167,67$
quindi non molto differente da quello ricavato prima.

PROPRIETA' DEI VARI TIPI DI MEDIA

Se i nostri dati sono $x_1, x_2, \dots, x_n$, e la loro media aritmetica è $M = \dfrac{x_1 + x_2 + \dots + x_n}{n}$,

i loro "scarti" dalla media sono le differenze fra i dati stessi e la media:

$$\left.\begin{matrix} x_1 - M \\ x_2 - M \\ \dots \\ x_n - M \end{matrix}\right\} \textit{scarti dei dati dalla media}$$

Ecco la tabella delle altezze superate, in cm, da 7 atleti ad una gara dilettantistica di salto in alto:

180	180	184	184	184	190	200

Se ne calcoli la media, avrai

$$M = \frac{180+180+184+184+184+190+200}{7} = \frac{1302}{7} = 186$$

Ora scriviamo, sotto ciascuno dei dati, il suo scarto dalla media:

180	180	184	184	184	190	200
–6	–6	–2	–2	–2	+4	+14

Se a questo punto sommiamo algebricamente questi 7 scarti, avremo

$$-6-6-2-2-2+4+14=0$$

Il fatto che la somma algebrica degli scarti dalla media aritmetica sia 0 è del tutto generale.

In effetti, se $x_1, x_2, \dots, x_n$ sono i dati,
e quindi $x_1 - M, x_2 - M, \dots, x_n - M$ sono i loro scarti dalla media aritmetica, avremo

$$\textit{somma scarti} = (x_1 - M) + (x_2 - M) + \dots + (x_n - M) = x_1 - M + x_2 - M + \dots + x_n - M =$$

$$= x_1 + x_2 + \dots + x_n - nM = x_1 + x_2 + \dots + x_n - \cancel{n} \cdot \frac{x_1 + x_2 + \dots + x_n}{\cancel{n}} = x_1 + x_2 + \dots + x_n - (x_1 + x_2 + \dots + x_n) = 0$$

PROPRIETÀ: **La somma degli scarti dei dati dalla media aritmetica dei dati stessi è sempre uguale a 0.**

Un'altra proprietà interessante della media aritmetica è la seguente:

PROPRIETÀ (che non dimostriamo; potresti però verificarla su di un esempio, tramite un foglio elettronico …)
La media aritmetica è quel valore rispetto al quale è minima la somma dei QUADRATI degli scarti.
Vale a dire, se io calcolo la somma dei quadrati degli scarti dalla media aritmetica,
questa somma sarà certamente minore di ciò che otterrei se, al posto degli scarti dalla media aritmetica M,
considerassi gli scarti da un qualsiasi altro valore a.
Schematicamente: se $x_1, x_2, \dots, x_n$ sono i dati, e M è la loro media aritmetica, allora la quantità

$$(x_1 - a)^2 + (x_2 - a)^2 + \dots + (x_n - a)^2$$

è minima nel caso $a = M$.

RIASSUNTO SCHEMATICO (INDICI DI POSIZIONE: le medie "ferme")

$\boldsymbol{\textit{MEDIA ARITMETICA} = M = \dfrac{x_1 + x_2 + \dots + x_n}{n}}$	"Conserva" la somma. EXCEL, OPENOFFICE CALC: = **MEDIA()**
$\boldsymbol{\textit{MEDIA GEOMETRICA} = M_G = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}}$	"Conserva" il prodotto
$\boldsymbol{\textit{MEDIA QUADRATICA} = M_Q = \sqrt{\dfrac{x_1^2 + x_2^2 + \dots + x_n^2}{n}}}$	"Conserva" la somma dei quadrati
$\boldsymbol{\textit{MEDIA ARMONICA} = M_A = \dfrac{1}{\dfrac{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}{n}}}$	"Conserva" la somma dei reciproci

B) MEDIE PONDERATE

1	9
2	14
3	25
4	22
5	10
	80

Avevamo già fatto qualche anticipazione. Riprendiamo il discorso.
Un test su 80 studenti universitari ha fatto registrare i punteggi della tabella qui a fianco (accanto a ciascun possibile punteggio, da 1 a 5, è stata annotata la relativa "frequenza", ossia il numero di studenti che hanno conseguito *quel* punteggio).
Qual è stata la media dei punteggi di questo gruppo di studenti?

$$M=\frac{\overbrace{1+...+1}^{9\ addendi}+\overbrace{2+...+2}^{14\ addendi}+\overbrace{3+...+3}^{25\ addendi}+\overbrace{4+...+4}^{22\ addendi}+\overbrace{5+...+5}^{10\ addendi}}{80}=\boxed{\frac{1\cdot 9+2\cdot 14+3\cdot 25+4\cdot 22+5\cdot 10}{80}}=$$

$$=\frac{9+28+75+88+50}{80}=\frac{250}{80}=3{,}125$$

Si dice "MEDIA PONDERATA" (o "MEDIA PESATA") una media nella quale ciascun dato viene moltiplicato per un fattore dato dalla sua frequenza assoluta, ossia è contato per un numero di volte uguale alla sua frequenza assoluta.
Nell'esempio di cui sopra, il dato "1" ha "peso" 9, il dato "2" ha "peso" 14, ecc.

Dunque, in generale, si ha,
per una media (aritmetica) ponderata,

$$\boxed{\begin{array}{l}\textit{Media}\\ \textit{aritmetica}=M=\dfrac{x_1\cdot f_1+x_2\cdot f_2+...+x_n\cdot f_n}{\underbrace{f_1+f_2+...+f_n}_{numero\ totale\ dei\ dati}}\\ \textit{ponderata}\end{array}}=\frac{\sum_{k=1}^{n}x_k\cdot f_k}{\underbrace{\sum_{k=1}^{n}f_k}_{numero\ totale\ dei\ dati}}$$

f_k = *frequenza assoluta del* k*-esimo dato*
(*numero di volte in cui compare*)

In questa interpretazione (ma vedi poi il riquadro sottostante), una "media ponderata" non differisce da una normalissima media. Semplicemente, visto che un dato si è presentato nella rilevazione più volte, lo si scrive, per comodità, una volta sola, *moltiplicandolo per la sua frequenza*, ossia per il numero di volte in cui compare.

Il simbolo $\sum_{k=...}^{...}$ si chiama "simbolo di *sommatoria*".

Scrivere $\sum_{k=1}^{n}x_k\cdot f_k$ significa che si vuole eseguire la somma di tanti addendi $x_k\cdot f_k$, dove k assume:
- il valore 1 (1° addendo),
- poi il valore 2 (2° addendo),
- eccetera,
- fino al valore n.

❑ Verifica che se 26 persone hanno donato 5 euro e 14 persone 10 euro, la media dell'offerta è stata di euro 6,75

❑ Calcola la media del voto in condotta, nella pagella più recente, di tutti gli studenti della tua classe

Si parla di "media ponderata" anche quando si vogliono assegnare, ai dati $x_1, x_2, ..., x_n$, "PESI" DIVERSI in quanto i dati vengono ritenuti di diversa "importanza". La formula è la stessa, solo che al posto delle frequenze $f_1, f_2, ..., f_n$ ci sono i "pesi" $p_1, p_2, ..., p_n$ che si è deciso di attribuire ai vari dati; vedi qui a destra →

Un'ALTRA INTERPRETAZIONE
♥ della media ponderata

$$\boxed{M=\frac{x_1\cdot p_1+x_2\cdot p_2+...+x_n\cdot p_n}{\underbrace{p_1+p_2+...+p_n}_{somma\ dei\ pesi}}}$$

❑ Verifica che se i punteggi ottenuti da uno studente per i tre esercizi A, B, C sono stati rispettivamente 10, 8 e 7, e ai tre esercizi l'insegnante ha ritenuto di attribuire rispettivamente i "pesi" 2, 3 e 3, allora il punteggio dato dalla media ponderata risulta essere 8,125.
Verifica poi che se invece i "pesi" fossero 2, 3 e 4, quello studente otterrebbe come punteggio finale 8, mentre la media aritmetica "semplice" (= non ponderata) degli stessi punteggi è 8,333…

❑ Se in una doppia prova scritto+orale Anna è stata valutata rispettivamente 7 e 9, e la prof. intende assegnare peso 3 allo scritto e peso 1 all'orale per fare poi una media ponderata tramite la formula nell'ultimo riquadro, è come se Anna avesse preso i 4 voti 7, 7, 7 e 9 di cui fare poi la media "semplice" (= normale): verificalo.

C) LE MEDIE "LASCHE": MEDIANA E MODA
(lasco = *allentato, molle, non teso*)

(Una media si dice "lasca" se potrebbe pure restare invariata, qualora cambiasse uno dei termini)

4	1
5	5
6	7
7	7
8	3
9	3
10	1
	27

In un test, i punteggi dei 27 studenti sono stati quelli riassunti dalla tabella qui a fianco (punteggio sulla colonna sinistra, frequenza assoluta di quel punteggio sulla colonna destra).
Se trascriviamo i punteggi uno a uno in ordine crescente, avremo

4	**5**	**5**	**5**	**5**	**5**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**7**	**7**	**7**	**7**	**7**	**7**	**7**	**8**	**8**	**8**	**9**	**9**	**9**	**10**

Consideriamo ora il punteggio che, nella striscia, occupa la **posizione centrale**:

4	**5**	**5**	**5**	**5**	**5**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**7**	**7**	**7**	**7**	**7**	**7**	**7**	**8**	**8**	**8**	**9**	**9**	**9**	**10**

Questo punteggio è 7. Diremo dunque che la "**mediana**" della distribuzione in esame è 7.
Se poi il "mostro" che ha preso 10 fosse stato assente, la striscia dei punteggi avrebbe contenuto 26 numeri:

4	**5**	**5**	**5**	**5**	**5**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**7**	**7**	**7**	**7**	**7**	**7**	**7**	**8**	**8**	**8**	**9**	**9**	**9**

In questo caso un numero che occupi la posizione centrale … non c'è. In tale eventualità (**numero pari di dati**) **si assume, convenzionalmente, come mediana la semisomma (= la media aritmetica) dei due valori che stanno all'immediata sinistra e all'immediata destra della posizione centrale**.

Nell'esempio considerato, quindi, la *mediana* sarebbe stata $\frac{6+7}{2}=6,5$.

Osserviamo che la *media aritmetica* dei punteggi della classe è
$\frac{181}{27}\approx 6,704$ con la presenza del "mostro", $\frac{171}{26}\approx 6,577$ supponendo il "mostro" assente. Dunque:

La **MEDIANA** è definita quando si ha un insieme di dati, disposti in ordine crescente.
Si tratta allora del **dato che "occupa il posto centrale"** ,
nel senso che **metà dei dati considerati sta a sinistra e metà a destra della mediana**.

Nel caso in cui il numero di questi dati sia pari, un "dato centrale" non esiste
e quindi, convenzionalmente, si assume come mediana la media aritmetica
fra i due dati che stanno immediatamente prima e immediatamente dopo, rispetto alla posizione centrale.

Qualora i dati non siano numerici, ma abbia comunque senso ordinarli (livelli di istruzione, aggettivi che esprimono un gradimento …) non ha senso pensare ad una "media" … ma a una *mediana*, in generale, sì.

Quando è possibile determinare sia la media aritmetica che la mediana, cioè con dati numerici, la "mediana" ci dà un'informazione diversa rispetto alla "media".

Abbiamo già visto che la media aritmetica è quel valore che,
se venisse sostituito al posto di ciascuno di dati, ne lascerebbe inalterata la somma;
la mediana ci dice invece qual è il valore "centrale" della successione di dati, nel senso che,
se conosciamo la mediana, possiamo dire che un 50% dei dati è $\le$ e l'altro 50% è $\ge$ della mediana.

La mediana, rispetto alla media aritmetica, è meno "sensibile" alla presenza di "dati anomali", cioè di dati "lontani dalla centralità".

Se nel precedente insieme di punteggi il punteggio più basso fosse stato "2" anziché "4", la mediana non sarebbe variata, la media aritmetica sì.

PROPRIETÀ: **La mediana, se è un valore numerico, è quel valore rispetto al quale è minima la somma dei valori assoluti degli scarti.**
Vale a dire, se io calcolo la somma dei valori assoluti degli scarti dalla mediana,
questa somma sarà certamente minore di ciò che otterrei se, al posto degli scarti dalla mediana,
considerassi gli scarti da un qualsiasi altro valore. *Verificalo empiricamente col foglio elettronico*!

E parliamo, infine, di **moda**. 

Per MODA si intende **il dato che si è presentato con più frequenza**.

La moda potrebbe anche non essere unica!

Nell'esempio sopra considerato della classe col suo test,
ci sarebbero state due "mode": 6 e 7 (si parla in questo caso di distribuzione "bimodale").

punt	freq
4	1
5	5
6	**7**
7	**7**
8	3
9	3
10	1

Nel caso in cui le modalità sono suddivise in "classi", più che parlare di "moda"
è corretto parlare di "**CLASSE MODALE**" (= la classe con maggiore frequenza).

Quando abbia senso parlare tanto di media aritmetica, quanto di mediana, quanto di moda, la moda ci dà un'informazione diversa rispetto alla media e alla mediana.

E osserviamo che nel caso in cui i dati siano di carattere qualitativo, e non abbia gran senso ordinarli, non si può parlare né di media né di mediana, mentre la moda è comunque determinabile.

Se ad esempio un certo giorno di Agosto una gelateria ha venduto
12 granite al limone, 15 all'arancia e 7 al cedro, quel giorno la "moda" per le granite è stata "arancia",
senza che ovviamente si potesse parlare né di media aritmetica né di mediana.

UN'ESERCITAZIONE COL FOGLIO ELETTRONICO: MEDIE, CONTEGGI, ISTOGRAMMA

I pesi in kg dei 240 maschi maggiorenni di un villaggio sul fiume Yukon, in Alaska, sono stati registrati in un foglio elettronico.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	70,0	80,5	80,7	83,9	72,7	82,8	83,6	69,5	73,8	78,6	73,1	101,2	60,6	80,4	76,4	76,5
2	78,6	69,1	93,2	69,4	76,2	61,9	101,8	104,3	88,8	75,8	77,7	94,7	67,4	79,9	77,5	86,9
3	69,9	87,2	83,2	110,8	99,1	80,4	85,6	73,4	94,9	72,5	74,3	75,3	61,6	102,9	83,1	99,4
4	83,8	75,0	68,7	87,8	112,2	68,6	73,7	64,5	83,3	85,3	68,2	88,5	57,8	65,9	80,9	70,4
5	83,6	77,9	70,2	101,9	87,0	88,9	71,0	81,5	96,0	70,8	86,3	72,8	71,8	68,5	73,9	86,1
6	95,4	77,7	70,4	73,8	91,7	83,6	89,4	57,4	81,2	94,6	77,5	72,5	63,2	109,4	79,5	57,8
7	82,1	89,4	71,1	81,6	89,2	63,5	90,0	76,9	90,9	93,7	76,2	63,7	62,3	84,9	71,7	101,8
8	79,8	71,2	76,0	70,9	114,7	99,2	78,8	90,0	63,6	65,2	75,8	98,1	69,3	106,5	80,4	106,3
9	86,6	76,3	66,6	76,2	92,1	98,4	78,4	79,2	67,5	101,2	71,6	76,3	61,8	99,5	81,2	103,3
10	89,9	84,4	72,9	75,9	119,2	75,4	89,2	76,1	68,6	69,1	72,6	88,3	89,8	53,8	86,6	90,5
11	84,5	75,6	56,7	77,5	93,0	101,9	80,3	67,0	72,2	109,7	80,2	78,4	82,3	66,1	85,1	70,5
12	98,0	85,2	64,9	80,9	98,4	103,5	75,1	82,7	59,6	66,2	79,8	99,3	91,3	72,2	93,7	97,8
13	81,9	76,3	67,2	68,0	96,2	78,4	90,8	79,6	67,1	71,1	80,3	67,7	91,9	77,3	84,0	60,9
14	93,1	96,5	73,7	90,9	56,8	69,1	92,2	73,4	60,4	90,8	81,5	70,1	81,5	76,0	72,7	91,2
15	85,2	80,2	80,6	83,5	74,5	57,0	88,0	71,6	72,9	77,8	75,0	90,4	98,0	105,6	68,7	84,7

a) Determinare il peso minimo e il peso massimo
b) determinare media e mediana dei pesi
c) contare il numero di persone il cui peso rientra nella fascia da 50 kg a 60 kg (esclusi), da 60 a 70, ecc., e tracciare un istogramma;
d) determinare la media dei pesi suddivisi in "classi", assegnando a ogni classe il peso centrale fra i suoi due estremi, e ricalcolare la media per confrontarla con la media reale.

a) Possiamo posizionarci in una cella libera qualsiasi, ad esempio la A18, e digitare

= min(A1 : P15)

> **BELLO**
> Osserviamo che dopo aver digitato **= min(**
> se **clicchiamo sulla cella A1** il foglio inserirà **automaticamente** nella formula il riferimento ad A1.
> A questo punto digiteremo i "due punti": dopodiché potremo cliccare su P15 e infine chiudere la parentesi.
> **Comoda alternativa**: si può digitare **= min(** poi **TRASCINARE** il mouse sul rettangolo da A1 a P15.

L'effetto finale, in A18, sarà

17	
18	53,8
19	

Allo stesso modo, in B18 inseriremo la formula

= max(A1 : P15)

ottenendo

17		
18	53,8	119,2
19		

Naturalmente, sarà opportuno inserire in celle adiacenti, **stringhe adeguate** che ci aiutino a ricordare il significato dei numeri ottenuti: ad es.

18	53,8	119,2
19	min	max

b) Digitiamo, ad esempio in C18 e in D18, le formule

= media(A1 : P15)

e rispettivamente

= mediana(A1 : P15) …

… nonché, in C19 e D19, le **stringhe opportune**, con l'effetto seguente:

18	53,8	119,2	80,8	79,7
19	min	max	media	mediana

c) Digitiamo in E18:

= conta.se(A1 : P15;" >= 50") – conta.se(A1 : P15;" >= 60")

e ci comparirà così, in E18, il numero di dati compresi fra 50 (incluso) e 60 (escluso):

	A	B	C	D	E
17					50-60
18	53,8	119,2	80,8	79,7	8
19	min	max	media	mediana	

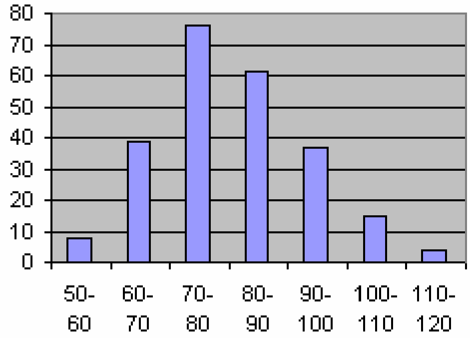
Procediamo in modo analogo sulle celle F18, G 18 … fino a K18:

	A	B	C	D	E	F	G	H	I	J	K
17					50-60	60-70	70-80	80-90	90-100	100-110	110-120
18	53,8	119,2	80,8	79,7	8	38	77	61	37	15	4
19	min	max	media	mediana							

Ora possiamo **selezionare, trascinando col mouse, il rettangolo di celle E17:K18**

17					50-60	60-70	70-80	80-90	90-100	100-110	110-120
18	53,8	119,2	80,8	79,7	8	38	77	61	37	15	4
19	min	max	media	mediana							

… **cliccare su** e, con qualche passaggio molto intuitivo, **ottenere finalmente l'istogramma**:



d) **Digitiamo, accanto alle frequenze delle classi, il "valore centrale" della classe …**

17					50-60	60-70	70-80	80-90	90-100	100-110	110-120
18	53,8	119,2	80,8	79,7	8	38	77	61	37	15	4
19	min	max	media	mediana	55	65	75	85	95	105	115

… e avviamoci ora a calcolare una MEDIA PONDERATA. In E20 inseriremo la formula
= E18 * E19

E20 =E18*E19

	A	B	C	D	E	F	G	H	I	J	K
16											
17					50-60	60-70	70-80	80-90	90-100	100-110	110-120
18	53,8	119,2	80,8	79,7	8	38	77	61	37	15	4
19	min	max	media	mediana	55	65	75	85	95	105	115
20					440						

che poi incolleremo, trascinando il quadratino in basso a destra della cella, sulle celle limitrofe F20 … K20

17					50-60	60-70	70-80	80-90	90-100	100-110	110-120
18	53,8	119,2	80,8	79,7	8	38	77	61	37	15	4
19	min	max	media	mediana	55	65	75	85	95	105	115
20					440	2470	5775	5185	3515	1575	460

Ora in L18 e in L20 calcoliamo la **somma delle frequenze assolute** e, risp., la **somma dei prodotti** …

17					50-60	60-70	70-80	80-90	90-100	100-110	110-120	
18	53,8	119,2	80,8	79,7	8	38	77	61	37	15	4	240
19	min	max	media	mediana	55	65	75	85	95	105	115	
20					440	2470	5775	5185	3515	1575	460	19420

… per terminare in bellezza con la formula, inserita in L21:
= L20 / L18
che ci dà la "**MEDIA PER CLASSI**", molto vicina, come possiamo osservare, alla vera media precedentemente determinata.

15	4	240
105	115	
1575	460	19420
		80,917