

MEDIOMAN

LUCA GRANIERI

Qualche tempo fa, il comico F. De Luigi proponeva un personaggio di nome: *Medioman*. Si trattava di una specie di *supereroe* che portava all'estremo i difetti e i vizi più comuni, quelli dell'uomo *medio* insomma. Anche per questo risultava divertente, poiché chiunque poteva in parte riconoscersi. Il fatto è che, tra tutte le possibilità, riteniamo la *media* come la più rappresentativa. Questo lo sanno molto bene ad esempio gli studenti. Uno studente è infatti sottoposto a varie valutazioni e in genere alla fine il suo *valore* è misurato appunto dalla media di tutte le sue valutazioni. Perché si prende proprio la media? Su cosa si basa la nostra fiducia in questo particolare valore?

1. DISPERSIONE

Intanto osserviamo che la rappresentatività della media dipende anche da quanto i dati sono *dispersi*. Consideriamo uno studente che realizzi le votazioni:

3; 3; 8; 3; 8; 9

La media di questi voti è $M_1 = \frac{3+3+9+3+9+9}{6} = 6$. Un secondo studente invece magari ottiene le votazioni:

5; 7; 6; 6; 7; 5

la media corrispondente è $M_2 = \frac{5+7+6+6+7+5}{6} = 6$. Dunque, entrambi gli studenti verranno giudicati con una valutazione complessiva di *sufficiente*. Ma c'è qualcosa di strano. Mentre per il secondo studente siamo abbastanza sicuri che il suo valore sia grosso modo *sufficiente*, per il primo c'è qualcosa che non quadra, sembrando oscillare paurosamente tra l'eccellenza e la disfatta totale. Come distinguere le due situazioni? A tal fine è utile introdurre la nozione di **Varianza o Scarto quadratico medio**, denotata usualmente con σ^2 , come la media aritmetica della somma degli scarti al quadrato. Ovvero

$$\sigma^2 = \frac{(x_1 - M)^2 + (x_2 - M)^2 + \dots + (x_n - M)^2}{n}.$$

Dove gli x_i denotano i dati a disposizione (nel nostro esempio i voti degli studenti), mentre M è la media aritmetica dei dati stessi.

La varianza non è mai negativa, essendo somma di quantità positive. Più grande è la varianza più i dati sono dispersi rispetto alla media M , più è piccola più i dati sono concentrati attorno alla media. Nel caso dei nostri due studenti, Valutando la varianza si ottiene

$$\begin{aligned}\sigma_1^2 &= \frac{(6-3)^2 + (6-3)^2 + (6-8)^2 + (6-3)^2 + (6-8)^2 + (6-9)^2}{6} = \\ &= \frac{9+4+9+4+9+9}{6} = \frac{44}{6}; \\ \sigma_2^2 &= \frac{(6-5)^2 + (6-7)^2 + (6-6)^2 + (6-6)^2 + (6-7)^2 + (6-5)^2}{6} =\end{aligned}$$

$$\frac{1 + 1 + 0 + 0 + 1 + 1}{6} = \frac{4}{6}.$$

La differenza tra le due situazioni è che la varianza del primo studente è sensibilmente più grande rispetto a quella del secondo (più di dieci volte). Questo aspetto può essere reso, come vedremo tra non molto, più quantitativo (e quindi utile). Nel caso specifico, i voti del primo studente sono piuttosto dispersi rispetto al valore medio, mentre per il secondo questi sono molto più *raggruppati* attorno al valor medio. Allora, per il primo studente un'indagine ulteriore è d'obbligo e magari scopriremo che ha qualche *asso nella manica* oppure qualche problematica che disturba le sue performance in certe circostanze. Risulta utile definire anche la cosiddetta **deviazione standard**, definita tramite la media aritmetica semplice del quadrato degli scarti. Ovvero

$$\sigma = \sqrt{\frac{(x_1 - M)^2 + (x_2 - M)^2 + \cdots + (x_n - M)^2}{n}}.$$

Le motivazioni sono diverse. Intanto, la deviazione standard ha la stessa unità di misura della media e quindi dei dati di partenza. Non meno importante è il fatto che la deviazione standard rappresenta una distanza (come quella tra due punti nel piano cartesiano) media (di tipo euclideo) tra i dati e il valor medio. Quindi, se vogliamo, rappresenta una *distanza* tra i dati considerati nel loro complesso e il valor medio.

La media soddisfa le seguenti importanti proprietà

Teorema (Proprietà della media aritmetica). *Dati n valori x_i , detta M la media aritmetica si ha che*

- M è compresa tra il più piccolo e il più grande tra i valori x_i ;
- (Scarto semplice) $\sum_{i=1}^n (x_i - M) = 0$;
- La media minimizza lo scarto quadratico (e quindi la deviazione standard)

Dimostrazione. Ordiniamo se necessario gli n valori x_i . Cambiando se serve gli indici possiamo supporre che sia $x_1 \leq x_2 \leq \cdots \leq x_n$. Sommando si ottiene $n \cdot x_1 \leq \sum_{i=1}^n x_i \leq n \cdot x_n$. Dividendo per n si ottiene $x_1 \leq M \leq x_n$.

Per quanto riguarda lo scarto semplice dei valori basta osservare

$$\sum_{i=1}^n (x_i - M) = \sum_{i=1}^n x_i - n \cdot M = n \cdot M - n \cdot M = 0.$$

Infine, per valutare che la media realizza la migliore concentrazione dei dati, denotiamo con $f(x) = \sum_{i=1}^n (x_i - x)^2$. Scrivendo $f(x) = f(x - M + M)$ e sviluppando il quadrato del binomio si ottiene

$$\begin{aligned} f(x) &= f(x - M + M) = \sum_{i=1}^n (x_i - M + M - x)^2 = \\ &= \sum_{i=1}^n (x_i - M)^2 + \sum_{i=1}^n (M - x)^2 + \sum_{i=1}^n 2(x_i - M)(M - x) = \\ &= f(M) + \sum_{i=1}^n (M - x)^2 + 2(M - x) \sum_{i=1}^n (x_i - M) = f(M) + \sum_{i=1}^n (M - x)^2 \end{aligned}$$

dove si è tenuto conto che $\sum_{i=1}^n (x_i - M) = 0$. Segue allora che $f(M) \leq f(x)$. \square

Prima di concludere la sezione, precisiamo che la formula per la deviazione standard è leggermente diversa da quella usualmente utilizzata. Ci sono infatti ragioni teoriche e pratiche per sostituire la formula precedente con la seguente

$$(1) \quad \sigma = \sqrt{\frac{\sum_{i=1}^n (M - x_i)^2}{n - 1}}.$$

In pratica, abbiamo sostituito il denominatore n con $n - 1$. Uno dei motivi, anche se non il più importante, è che nel caso limite di un solo dato a disposizione, la media coincide ovviamente con il dato e la varianza è nulla. Allora dovremmo concludere erroneamente che il dato ha una concentrazione massima. Ma nessuno studente vorrebbe che il suo lavoro di un anno fosse valutato sulla base di una sola prova! La (1) evita (non può essere applicata per $n = 1$) il caso paradossale di un solo dato.

2. DISTRIBUZIONI GAUSSIANE

Consideriamo ora una serie di dati soggetti a piccole variazioni casuali, come possono essere le misure delle altezze di una certa popolazione, o delle misure in un generico esperimento di fisica. E rappresentiamo in un istogramma di frequenza i dati a disposizione. Aumentando il numero di dati, l'istogramma conseguente tende ad assumere una forma geometrica sempre più definita, approssimandosi ad una curva a *campana*

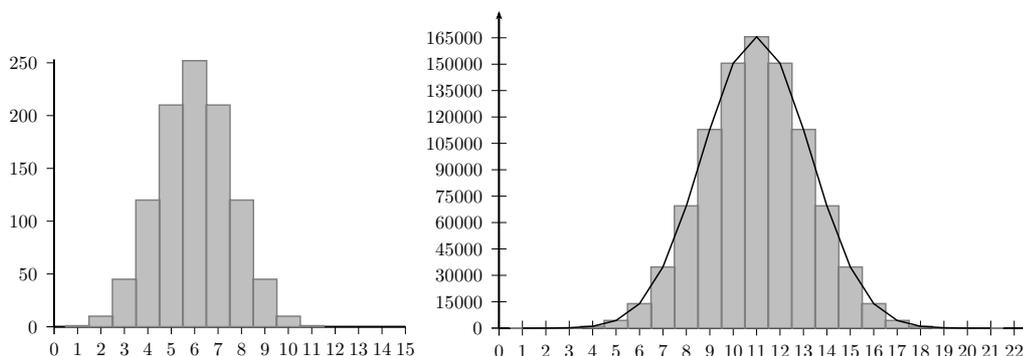


FIGURA 1. Istogrammi di frequenza.

Si dimostra che la distribuzione di questi dati si accorda (per grandi numeri di dati) ad un grafico a *campana* detto distribuzione Gaussiana (o *normale*). E la funzione che regola la distribuzione di questi dati è del tipo:

$$(2) \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-M)^2}{2\sigma^2}}.$$

In questa formula, M corrisponde all'asse di simmetria della distribuzione e al suo punto di massimo. Mentre σ quantifica la dispersione dei dati attorno ad M , distinguendo tra distribuzioni *basse e tozze* oppure *alte e slanciate*. Nella (2) π è essenziale e rende la distribuzione di *probabilità*, assicurando cioè che l'area racchiusa dal grafico sia pari ad uno. Così, la probabilità di osservare un dato x compreso tra due valori x_1 e x_2 è dato dalla area racchiusa dalla campana tra x_1 e x_2 . Se aumentiamo σ la campana si *allarga* diventando automaticamente più *bassa*. Se invece

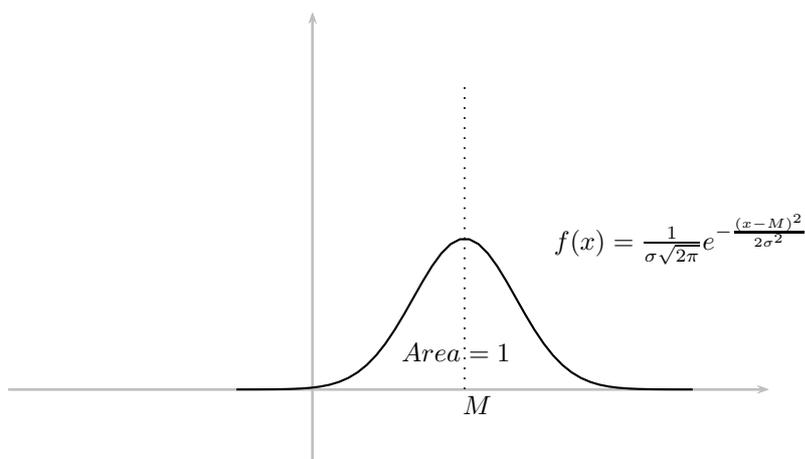


FIGURA 2. Distribuzione di Gauss

diminuiamo σ la campana si *stringe* diventando automaticamente più alta. Come mai? Perché l'area sottesa deve essere sempre pari ad uno (fig. 3), trattandosi di una distribuzione di probabilità.

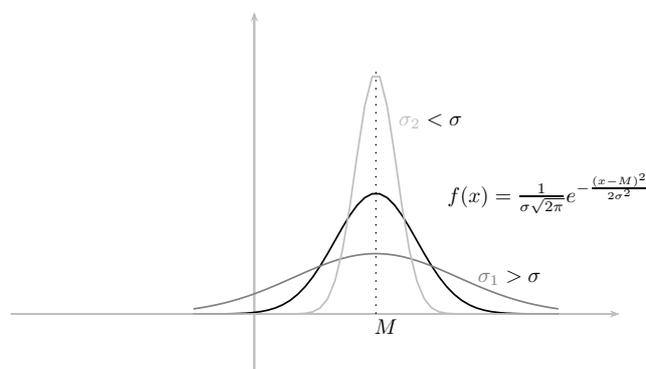


FIGURA 3. Comportamento della gaussiana rispetto alla deviazione standard.

Modificare invece M corrisponde a cambiare il centro della campana, trasladola a destra o a sinistra. Aver denotato con M, σ i parametri è, come vedremo, non casuale, è proprio il caso di dirlo. Comunque sia, si tratta di una di quelle formule che farebbe la sua figura stampata su una maglietta, o se non sapete che tatuaggio farvi dietro la schiena.

Eccola in tutto il suo splendore (purtroppo stampata in piccolo vicino al grafico a campana) su una banconota tedesca.



Banconota commemorativa in onore di Gauss.

La distribuzione gaussiana si osserva tutte le volte che le variazioni di un carattere sono dovute ad un gran numero di cause indipendenti, producendo effetti molto piccoli e dello stesso ordine di grandezza. Tali sono per esempio i caratteri biometrici (altezza, peso, pressione sanguigna, ecc.), le variazioni metereologiche,

gli errori sperimentali nella misurazione di una grandezza fisica, la lunghezza dei bulloni prodotti da una macchina, la gittata di un proiettile, ecc. La distribuzione gaussiana è molto importante. Ad esempio per giustificare il fatto che la media M sia un buon candidato per esprimere la *misura* di riferimento relativa a delle misure x_i . Si può dimostrare che date le misure x_i , se queste si accordano con una distribuzione gaussiana (come è il caso di dati affetti da piccole variazioni casuali), allora la media M è quel valore che rende massima la probabilità di osservare proprio i dati x_i . Un discorso analogo può essere fatto per la deviazione standard σ . Quindi, i parametri M e σ sono in qualche modo i *migliori possibili*, secondo quello che si chiama *criterio della massima verosimiglianza*

3. MASSIMA VEROSIMIGLIANZA

Supponiamo che i nostri n dati x_i siano caratterizzati da una distribuzione di Gauss di parametri $\bar{M}, \bar{\sigma}$. La probabilità di osservare un valore tra $x_i - \varepsilon$ e $x_i + \varepsilon$ è data da

$$\mathcal{P}(x_i) = \frac{1}{\bar{\sigma}\sqrt{2\pi}} \int_{x_i-\varepsilon}^{x_i+\varepsilon} e^{-\frac{(x-\bar{M})^2}{2\bar{\sigma}^2}} dx = \frac{1}{\bar{\sigma}\sqrt{2\pi}} e^{-\frac{(c_i-\bar{M})^2}{2\bar{\sigma}^2}} \cdot (2\varepsilon)$$

per un certo $c_i \in [x_i - \varepsilon, x_i + \varepsilon]$ (per il teorema della media integrale).

Possiamo ragionevolmente ritenere che l'osservazione del singolo dato x_i sia indipendente dall'osservazione degli altri. In altre parole che si tratti di eventi indipendenti tra loro. Allora, la probabilità di osservare tutti gli n valori x_i corrisponde a

$$\mathcal{P}(x_1) \cdot \mathcal{P}(x_2) \cdots \mathcal{P}(x_n) = \frac{2^n \varepsilon^n}{\bar{\sigma}^n \sqrt{2^n \pi^n}} e^{-\sum_{i=1}^n \frac{(c_i - \bar{M})^2}{2\bar{\sigma}^2}}.$$

Se dunque abbiamo osservato gli x_i , l'idea è quella di cercare i valori di \bar{M} e di $\bar{\sigma}$ in modo che tale probabilità risulti la più grande possibile (massima verosimiglianza). Nel far questo, possiamo senz'altro trascurare i fattori costanti (che non dipendono né da \bar{M} e né da $\bar{\sigma}$). Si tratta dunque di massimizzare la funzione

$$(3) \quad f(\bar{M}, \bar{\sigma}) = \frac{1}{\bar{\sigma}^n} e^{-\sum_{i=1}^n \frac{(c_i - \bar{M})^2}{2\bar{\sigma}^2}}.$$

A tal fine, osserviamo che la dipendenza da \bar{M} non è complicata e già sappiamo che l'esponente dell'esponenziale è minimo proprio quando \bar{M} è la media aritmetica

dei c_i . Pertanto, a parità di $\bar{\sigma}$, la funzione f è massima per $\bar{M} = M_c$, dove l'indice ci ricorda che stiamo considerando la media dei valori c_i . Allora abbiamo che $f(\bar{M}, \bar{\sigma}) \leq f(M_c, \bar{\sigma})$. Ora, per $\varepsilon \rightarrow 0$ si ha che $M_c \rightarrow M$, con M la media degli x_i originari. Dunque, la media aritmetica è (arbitrariamente vicina alla) la miglior stima possibile.

A questo punto, un parametro è fissato, non ci resta che scegliere se possibile il restante parametro $\bar{\sigma}$ nel miglior modo possibile. Questa volta, la dipendenza da $\bar{\sigma}$ è più complicata. Il fatto è che se aumentiamo (risp. diminuiamo) $\bar{\sigma}$ la prima frazione nella (3) diminuisce (risp. aumenta) mentre l'esponenziale aumenta (risp. diminuisce). Occorre pertanto un'analisi più dettagliata. Avvelendoci dei rudimenti dell'*analisi matematica* valutiamo che

$$\lim_{\bar{\sigma} \rightarrow \pm\infty} f(M, \bar{\sigma}) = 0.$$

Come conseguenza del Teorema di Weierstrass, non è difficile verificare che allora la nostra funzione ha dei punti di massimo. Questi punti di massimo, come è noto, devono corrispondere a punti stazionari (critici) per f , ovvero punti con derivata nulla. Derivando (rispetto a $\bar{\sigma}$) nella (3) si ottiene

$$-\frac{n}{\sigma^{n+1}} e^{-\sum_{i=1}^n \frac{(c_i - \bar{M})^2}{2\bar{\sigma}^2}} + \frac{\sum_{i=1}^n (c_i - \bar{M})^2}{2\bar{\sigma}^{n+3}} e^{-\sum_{i=1}^n \frac{(c_i - \bar{M})^2}{2\bar{\sigma}^2}}$$

Pertanto, la derivata prima si annulla se e soltanto se

$$-\frac{n}{\sigma^{n+1}} + \frac{\sum_{i=1}^n (c_i - \bar{M})^2}{2\bar{\sigma}^{n+3}} = 0 \Leftrightarrow \bar{\sigma}^2 = \frac{\sum_{i=1}^n (c_i - \bar{M})^2}{n} \xrightarrow{\varepsilon \rightarrow 0} \frac{\sum_{i=1}^n (x_i - \bar{M})^2}{n}.$$

Pertanto, la miglior scelta possibile è proprio la deviazione standard σ . Alcune precisazioni. Intanto la media aritmetica e la deviazione standard degli n dati vanno considerate come *approssimazioni* dei parametri M, σ della distribuzione Gaussiana. Dunque, dati gli n dati, la media è la miglior approssimazione del cosiddetto *valore vero*, diciamo X . A causa delle proprietà della media che abbiamo dimostrato precedentemente, abbiamo che

$$\sum_{i=1}^n (x_i - \bar{M})^2 \leq \sum_{i=1}^n (x_i - X)^2.$$

Allora, la deviazione standard, come definita all'inizio, *sottistima* la dispersione *vera*. Questa osservazione è compatibile con la necessità di modificare l'espressione della deviazione standard rimpiazzando il denominatore con $n - 1$. Questa scelta è poi *la migliore* sotto altri punti di vista (gradi di libertà, correttezza statistica).

RIFERIMENTI BIBLIOGRAFICI

- [1] M. Dapor, M. Ropele, Elaborazione dei dati sperimentali, Springer 2005.
- [2] B. Giardina, Manuale di Statistica per aziende e ricercatori, F. Angeli, Milano 1962.
- [3] D. J. Hand, Il caso non esiste, BUR, Milano 2014
- [4] R. A. Johnson, Probabilità e statistica per Ingegneria e Scienze, Pearson, 2007.
- [5] D. A. Johnson, W. H. Glenn, M. Scott Norton, Caso e probabilità, Zanichelli, 1984.
- [6] J.G. Kemeny, J.L. Snell, G.L. Thompson, Matematica ed attività umane, Vol. 1, Feltrinelli.
- [7] L. Soliani, Manuale di Statistica per la ricerca e la professione, Aprile 2005 disponibile all'indirizzo: <http://www.dsa.unipr.it/soliani/soliani.html>
- [8] J. R. Taylor, Introduzione all'analisi degli errori, Zanichelli, 1990.