Misure di variabilità

1. Intervalli di variazione

I valori medi non sono sufficienti a caratterizzare una distribuzione rispetto ad un carattere quantitativo. Infatti, si possono avere distribuzioni molto differenti tra loro che presentano uno o più valori medi uguali.
Nell'esempio che segue, una classe di 20 alunni, suddivisa in due gruppi di 10 alunni, è stata sottoposta ad un test di ingresso che prevedeva 30 domande, con l'assegnazione di 1 punto per ciascuna risposta esatta e 0 punti per ciascuna risposta errata o non data. Ecco, nei due casi, le modalità del carattere "punteggio ottenuto al test":

a.	29	22	24	24	21	26	28	28	24	24
b.	24	23	27	29	24	19	21	30	24	29

Nel tentativo di sintetizzare i risultati del test l'insegnante ha calcolato il punteggio medio nei due casi, ottenendo:

$$m_a = \frac{21 + 22 + 24 \cdot 4 + 26 + 28 \cdot 2 + 29}{10} = 25$$

$$m_b = \frac{19 + 21 + 23 + 24 \cdot 3 + 27 + 29 \cdot 2 + 30}{10} = 25$$

25 compensa i voti alti e quelli bassi, ma ogni unità differisce da questo valore fittizio, non fosse altro perché, come punteggio, 25 non compare in alcuna delle due serie.
Si può facilmente verificare che, nei due casi, a fronte del fatto che il carattere studiato varia in modo differente, coincidono anche la moda (24) e la mediana (24). Allora come dare il senso di questa diversa variazione?

Diventa necessario trovare strumenti che esprimano in modo sintetico qual è l'attitudine del carattere studiato ad assumere modalità differenti, a variare. La variabilità, d'altronde, è una caratteristica della realtà. La ragion d'essere della statistica sta proprio nella variabilità dei fenomeni e nel conseguente tentativo di esprimerla e spiegarla.
Per i caratteri quantitativi, la più semplice misura di variabilità è la *differenza fra il valore massimo e il valore minimo* presentato dal carattere. Essa è indicata con V ed è chiamata **campo di variazione** ed è dotata della stessa unità di misura del carattere.
Nel caso precedente:

$$V_a = 29 - 21 = 8 \quad \text{e} \quad V_b = 30 - 19 = 11$$

e con questo risultato possiamo senz'altro evidenziare una prima differenza nella variabilità delle due serie.

Il campo di variazione tiene conto di tutti i dati, mentre la sua ampiezza V esprime la differenza massima fra due modalità osservate. V è sensibile alla eventuale presenza di osservazioni anomale (se un alunno del primo gruppo avesse riportato un punteggio molto basso, ad esempio 6, il campo di variazione sarebbe risultato molto dilatato, addirittura 23).

2. Scostamenti semplici medi

L'intervallo di variazione esprime la variabilità a partire dai due estremi osservati. Si può invece richiedere che la variabilità venga espressa utilizzando tutti i valori.
Rappresentando i punteggi della distribuzione a) come punti su di un segmento, sul quale viene indicata anche la media aritmetica, la diversità fra ciascun punteggio e la media m è indicata da un segmento (Fig. 1).

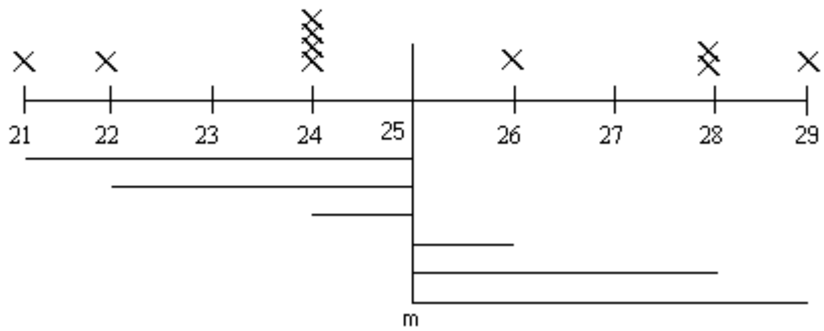

Fig. 1 - Distribuzione degli studenti secondo il punteggio riportato al test d'ingresso
e rappresentazione degli scostamenti dal punteggio medio

Chiamiamo **scostamento** di un valore osservato la sua *differenza positiva o negativa calcolata rispetto alla media aritmetica*.
Allora, visto che qualche dato *si scosta di più*, qualcun altro di meno possiamo pensare di trovare una misura dello scostamento calcolandolo come media degli scostamenti. Comprendiamo però subito che, essendo zero la somma delle differenze prese ciascuna col proprio segno, sarà zero anche la loro media (e questo per qualsiasi distribuzione). Infatti:

$$\frac{somma\ s\cos tamenti}{n^\circ\ valori} = 0$$

Anche nella distribuzione a) l'uguaglianza è facilmente verificabile:

$$\frac{(21-25)+(22-25)+(24-25)\cdot 4+(26-25)+(28-25)\cdot 2+(29-25)}{10} =$$

$$= \frac{(-4)+(-3)+(-1)\cdot 4+(+3)\cdot 2+(+4)}{10} = 0$$

Diventa così inutile ricorrere alla somma degli scostamenti e alla loro media. Quanto fatto, tuttavia, suggerisce l'idea di assumere non lo scostamento, ma il suo modulo come misura della distanza di un valore osservato dalla media.
Nasce così lo **scostamento semplice medio dalla media aritmetica (**S_μ) come *media dei valori assoluti degli scostamenti*.
Per la distribuzione dei punteggi a) di alunni avremo:

$$S_\mu = \frac{somma\ s\cos tamenti\ assoluti}{n^\circ valori} =$$

$$= \frac{|21-25|+|22-25|+|24-25|\cdot 4+|26-25|+|28-25|\cdot 2+|29-25|}{10} = \frac{4+3+1\cdot 4+1+3\cdot 2+4}{10} = 2{,}2$$

Questo risultato può essere letto in questo modo:

“la media aritmetica dei punteggi è 25, mediamente i punteggi stessi sono distanti, in più o in meno, dalla media 2,2 punti”.

Può essere interessante ora ripetere lo stesso procedimento per i punteggi degli alunni del gruppo b. Così facendo si ottiene come scostamento semplice medio dalla media aritmetica il valore 3 e come interpretazione del risultato:
“la media aritmetica dei punteggi è 25, mediamente i punteggi stessi sono distanti, in più o in meno, dalla media 3 punti”.
Si può concludere che il primo gruppo di alunni ha avuto un punteggio medio pari a quello del secondo gruppo, ma anche che mediamente i punteggi del primo gruppo sono meno dispersi attorno alla media stessa.

Ci si può chiedere: chi ha avuto 29 è più bravo se appartiene alla distribuzione a) o alla b)? Certo la differenza dalla media è sempre 4, ma fra 4/2,2 e 4/3 chi è più grande? Allora è meglio 29 in a) che in b). Infatti in a) 29 è il punteggio massimo!!!

Dal punto di vista geometrico lo scostamento è un segmento e si può costruire un intono della media di ampiezza pari al corrispondente scostamento semplice medio dalla media aritmetica.