

INTRODUZIONE ALLA STATISTICA DESCRITTIVA

Ciò che qui di seguito è presentato è conseguenza del lavoro (classe prima come esempio) di presa di coscienza dell'ambito statistico in cui emergono quotidianamente le informazioni, presa di coscienza che ha portato a discussioni in classe anche non formali, finché il problema spontaneamente ha richiesto di essere approfondito.

Ciò che presentiamo è una breve **sintesi individuale** di elaborazione. Il tipo di indagine non ha valore come contesto di dati particolarmente significativi, ma per l'attenzione data ad ogni elaborazione (non è documentata la discussione classe).

La metodologia di lavoro è comunque quella meglio esplicitata nel capitolo sulla probabilità.

Qui non si vede l'uso di strumenti tecnologici di elaborazione perché in una classe prima può parere utile lavorare prima manualmente. Si può in tal modo arrivare ad evidenziare proprio come il lavoro "manuale" da compiere possa essere eseguito da una macchina e del perché e del come questo avvenga (utilizzo di uno strumento attuale consono al lavoro da fare che naturalmente comporta diversa organizzazione pratica). Ovviamente questo potrebbe essere superato da precedenti attività in tale direzione. Ci pare però che l'uso di qualche strumento non cambi qui il contenuto logico e matematico del lavoro fatto.

Dall'approfondimento raggiunto con tali analisi del problema, si giunge facilmente al significato di sondaggio ed alla sua validità con l'attenzione da porre nella costruzione di un **campione significativo** che non avverrà mai nelle nostre indagini in modo da non poter generalizzare i risultati. Già nelle nostre prime indagini abbiamo immediatamente rilevato che, per esempio, è assurdo pensare di raccogliere un campione significativo delle famiglie italiane riguardo al numero dei figli, andando a fare interviste alle porte delle scuole. Come minimo mancherebbero le famiglie prive di figli. Lo stesso si potrebbe dire per interviste alle porte dei cinematografi sulle frequenze delle visioni di film ecc. oltre alla difficoltà intrinseca di valutare quale sia un sufficiente "numero" di dati da rilevare perché il sondaggio sia valido.

La stessa analisi critica è stata in seguito costantemente sfruttata per evidenziare il limite ed il valore di ogni indice e dell'impiego di grafici sostanzialmente corretti, ma dove la chiarezza dell'informazione può essere celata a chi non ne conosca bene la struttura. Per esempio se la moda o la media sono molto interessanti o importanti per una produzione industriale, non permettono da sole di capire la variabilità della popolazione, un grafico "tagliato" nella parte comune a tutte le strisce può gravemente falsare il significato delle differenze ad un "occhio" poco attento....

Nel triennio le esperienze si sono quotidianamente arricchite sia per l'ampiamiento del linguaggio col proseguire dell'attività matematica generale (razionali, percentuali, grafici di tipo diverso, legge empirica dei grandi numeri, ecc...), sia per l'uso della conoscenza stessa in attività interdisciplinari (molto interessanti le esperienze in tal senso di "Science across the World" un lavoro scientifico proposto dall'unione Europea alla scuola con unità didattiche relativamente a temi di interesse generale, con

collaborazione e scambi di risultati tra classi di Stati diversi. Ciò ha permesso agli allievi di confrontare abitudini e diversi modi di approfondire alcuni problemi scientifici che si legano alla vita quotidiana come la “raccolta rifiuti”), sia per l'uso dello strumento informatico. Riteniamo il “capitolo” della Statistica un esempio molto interessante di conoscenza sfruttata nel quotidiano.

ELABORAZIONE INDIVIDUALE

(per la classe prima)

di R. B.

La statistica è quella branca della matematica che ci aiuta nell'elaborare delle informazioni (dati), in modo chiaro, sintetico ed adatto ad essere facilmente interpretate.

Evidenziamo un'indagine di esempio.

a) DESIDERIAMO ANALIZZARE IL NUMERO DEI FIGLI NELLE FAMIGLIE DELLA NOSTRA CLASSE.

Richiediamo l'informazione "numero dei figli" (dato) direttamente a voce (se l'indagine fosse più vasta dovremmo naturalmente usare altri mezzi di raccolta di cui parleremo eventualmente in seguito) e sulla lavagna scriviamo le risposte:

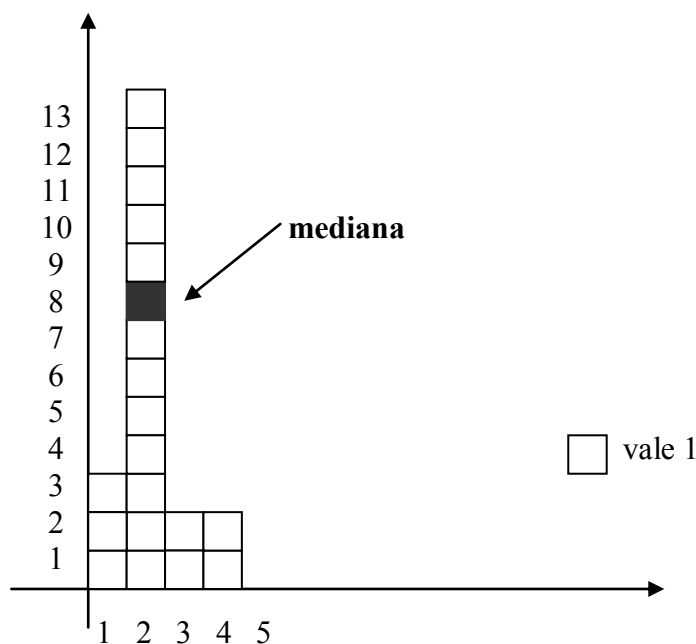
3,2,3,2,3,2,1,2,2,2,2,2,2,2,2,4,4,2,1,1,2

Se tutti hanno risposto il numero dei dati deve corrispondere al numero dei ragazzi della classe, cioè 21.

N. figli	1	2	3	4
N. famiglie	3	13	3	2

La tabella esplicita in modo chiaro ed immediato quante volte un certo dato si è verificato: diciamo **frequenza** (assoluta) il numero di volte con cui un dato si presenta.

Per esprimere ancora meglio, in modo da avere un'idea immediata della situazione, si può per esempio usare un disegno che associ alle frequenze dei "livelli" più o meno alti di strisce verticali: **i diagrammi a strisce** (tutte le strisce devono, naturalmente, partire dallo stesso livello).



Per essere corretto, il diagramma a strisce deve basarsi sulla stessa unità di misura e deve denunciare il livello di partenza uguale per tutti.

Riassumiamo le informazioni della **tabella e del grafico in:**

1. **moda** esprime il dato più frequente. Nel nostro caso la moda è il dato " 2 figli" con frequenza 13.
2. Un altro "riassunto" può essere dato dalla **media** di cui tutti sentiamo parlare e che sappiamo matematicamente calcolare sommando tutti i dati (che naturalmente devono essere numerici), che generalmente sono diversi l'uno dall'altro (ma non necessariamente), e dividendo la somma per il numero dei dati. **In questo caso la divisione riporta il discorso a come se le parti fossero tutte uguali. In pratica la divisione riporta l'equilibrio!**

Nel nostro caso:

$$m(\text{media}) = (3 * 1 + 13 * 2 + 3 * 3 + 2 * 4) : 21 =$$

$$= (3 * 26 + 9 + 8) : 21 = 46 : 21 = 2,1 \text{ troncato ai decimi.}$$

La media non è un numero naturale mentre i dati lo sono.

Ci verrebbe voglia di ritornare all'intero, ma ci rendiamo conto che nel comunicare la media col 2 o col 3 perdiamo in informazione,

Decidiamo allora di ignorare l'incongruenza e di accettare la media così come è per non perdere l'informazione del "più di 2" che essa contiene sapendo che **non è un dato reale, ma il risultato di un calcolo matematico.**

3. Un altro indice significativo, quando i dati sono numerici e quindi ordinabili in ordine crescente, è la **mediana**. Essa è **il dato di mezzo nella successione ordinata di dati**.

Come facciamo a ricavarla? Intuitivamente mi rendo conto che il lavoro varia se il numero dei dati è pari o dispari.

1° caso: numero dei dati dispari (anche il nostro, 21 dati).

$$M_e = \text{int} (21:2) + 1 = 10 + 1 = 11$$

La nostra **mediana** cade sull'undicesimo dato, 2figli.

In generale:

$$M_e = \text{int} (a : 2) + 1$$

2° caso: numero dei dati pari

In generale non avremo un dato centrale, ma una coppia di dati centrali

$$M_e = \begin{cases} n : 2 \\ n : 2 + 1 \end{cases}$$

dove n è il numero totale dei dati.

Perché la statistica ritiene importante questo indice? Perché dividendo il numero totale dei dati in due parti permette di evidenziare meglio il comportamento della popolazione. Nel nostro caso ci dice, per esempio, che metà delle famiglie della classe hanno da uno a due figli, l'altra metà ha due figli o più.

NOTA

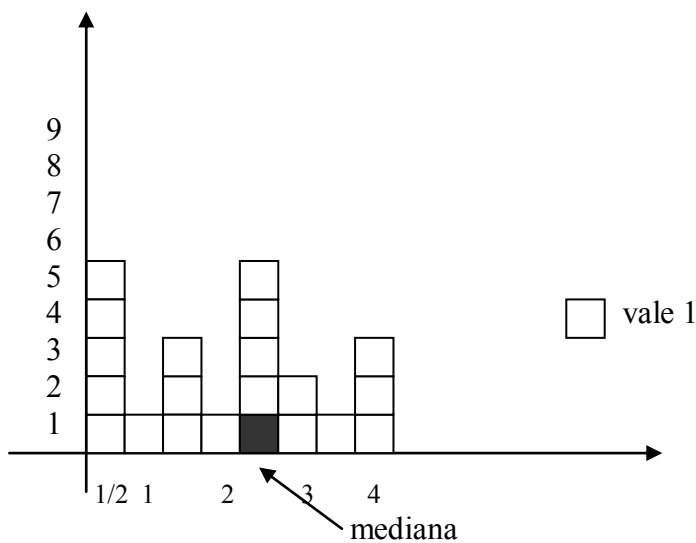
La **media**: questa informazione è abbastanza particolare. Non l'ho chiamata "dato" in quanto ritengo necessario precisare che l'assumiamo grazie ad un'elaborazione di tutti i dati numerici della nostra indagine.

Dall'esperienza precedente abbiamo notato che è molto utile, ma che può trarre in inganno. Dobbiamo pensare, quando ne entriamo in "possesso" che non è lo specchio della realtà, ma un'elaborazione di questa da parte della mente umana, che consente di farsi un'idea "approssimativa" dei fatti e di operare calcoli ed altre indagini, ma che non consente, da sola, di conoscere a fondo tutta la situazione. E' insomma una realtà apparente, da usare unicamente quando siamo costretti e, in ogni caso, quando siamo a conoscenza di tutta "l'impalcatura" che le sta dietro e possiamo controllare che sia coerente coi dati e in parte con la realtà.

b) OPERIAMO UN'ALTRA INDAGINE SULLE NOSTRE ORE DI STUDIO.

Dati	Raccolta	Frequenza
1/2 h	X X X X X	5
1 h	X	1
(1 + 1/2) h	X X X	3
2 h	X	1
(2 + 1/2) h	X X X X X	5
3 h	X X	2
(3 + 1/2) h	X	1
4 h	X X X	3
(4 + 1/2) h		0

Elaboro i dati:



Noto subito che i dati mezz'ora e due ore e mezza hanno la frequenza più alta, la stessa: non si potrà quindi parlare di moda, ma di due dati maggiori.

Per ricavare la media:

$$m = (5 * 0,5 + 1 * 1 + 1,5 * 3 + 2 * 1 + 2,5 * 5 + 3 * 2 + 3,5 * 1 + 4 * 3) : 21 =$$

$$= (2,5 + 1 + 4,5 + 2 + 12,5 + 6 + 3,5 + 12) : 21 = 44 : 21 =$$

44 : 21 = 2 ore col resto di 2 ore che trasformo in minuti per continuare la divisione

120 : 21 = 5 min col resto di 15 min che trasformo in sec per terminare il calcolo

900 : 21 = 42 sec con il resto di 18 sec. (qui potrei continuare con una numerazione decimale)

44:21 = 2 ore 5 minuti 42 secondi troncato ai secondi

Qui nasce però il problema dell'approssimazione da dare. Ci siamo infatti accorti che, anche le informazioni stesse erano state approssimate da chi le ha comunicate in quanto nessuno, credo, studi tutti i giorni 1 ora ecc. Ognuno ha operato forse una media sul numero di ora che studia in una settimana.. Anche nel richiedere i dati noi abbiamo tralasciato i minuti ferdandoci alla mezz'ora (arrotondata per difetto o per eccesso?). In questo caso il calcolo ci complica le cose. Dovremo approfondire il valore del risultato di un'indagine di questo tipo.

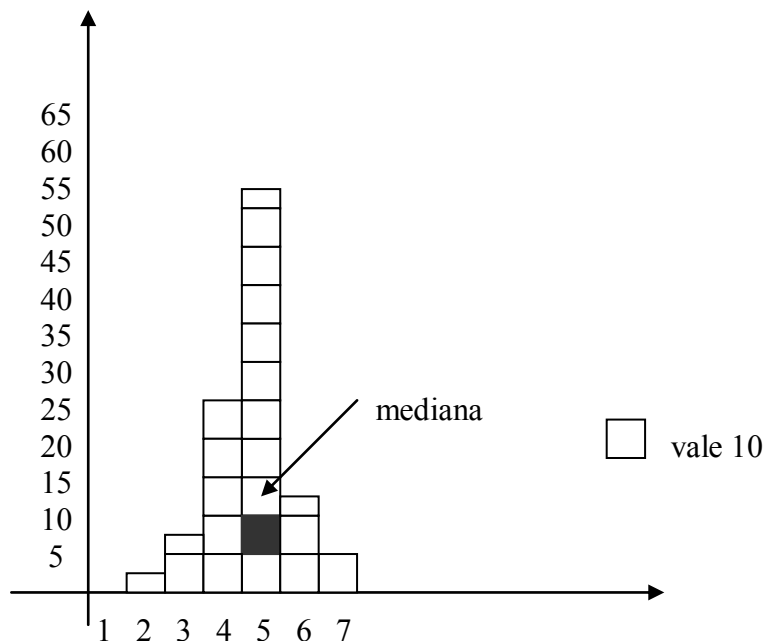
e) NUOVA INDAGINE SULLA RIPRODUZIONE DEGLI UCCELLI

Su di una rivista scientifica abbiamo notato la seguente tabella:

n. uova	2	3	4	5	6	7	
n. nidi	2	7	25	53	12	1	

Secondo me questa indagine è stata fatta per verificare il numero di uova nei nidi (di una particolare specie) magari in un particolare periodo, da studiosi.

Prima di elaborarli con calcoli, li chiarisco in un grafico:



Noto subito che la moda è rappresentata dal dato 5 uova, per la mediana, posso vederla direttamente sul grafico. Il numero delle uova è pari, avrei due risultati, ma coincidono con lo stesso dato , 5 uova.

Noto che il grafico presenta una certa simmetria rispetto alla moda. E' un dato di misura sugli esseri viventi, teniamone conto. **(Il discorso per ora lasciato come osservazione, è uno spunto che porterà poi alla curva di Gauss)**

$$M = (2 * 2 + 3 * 7 + 4 * 25 + 5 * 53 + 6 * 12 + 7 * 1) : 100 = \\ = 469 : 100 = 4,69$$

Questo indice, come nella prima indagine è incoerente con la realtà, ma come allora lo prendiamo come informazione dovuta ad un'elaborazione matematica. Essa permette di vedere in generale il comportamento della specie.