

APPROFONDIMENTO E SCHEDA OPERATIVA – ANALIZZARE E INTERPRETARE I DATI

Dopo aver raccolto i dati, il passo successivo è **analizzarli in modo riflessivo e contestualizzato**. Analizzare non significa solo “fare calcoli”, ma **osservare con metodo** ciò che emerge, riconoscere pattern e anomalie e iniziare a costruire ipotesi di interpretazione.

Questa fase è cruciale perché i dati, da soli, non parlano: **siamo noi a dare loro significato**, sulla base delle domande che ci poniamo, della nostra esperienza educativa e del confronto tra più punti di vista.

L’analisi deve essere graduale:

- **Descrivere:** guardare i dati “così come sono”, senza giudizio.
- **Stratificare:** analizzare per gruppi (es. genere, classi, bisogni educativi).
- **Rappresentare:** visualizzare ciò che si osserva (tabelle, grafici, mappe mentali...).
- **Interpretare:** ipotizzare cause, collegare fenomeni, formulare domande successive.

Un’analisi efficace prepara il terreno all’azione: più l’interpretazione è fondata, più le decisioni saranno pertinenti, sostenibili e condivise.

DOMANDE PER LA RIFLESSIONE

- Che cosa raccontano davvero questi dati sul problema osservato?
- Quali differenze emergono tra gruppi, tempi, contesti?
- Ci sono dati in contrasto con le nostre aspettative?
- Quali ipotesi ci portano a formulare questi dati? E su quali basi?
- Che altre informazioni ci servirebbero per completare il quadro?

DIFFERENZA TRA CORRELAZIONE E CAUSAZIONE

Per chiunque voglia analizzare rapporti statistici tra variabili è molto importante comprendere la differenza tra **correlazione** e **causazione**.

Due variabili possono apparire come correlate (variano insieme), ma questo non significa necessariamente che ci sia tra loro un legame di causazione, cioè che i due fenomeni siano l’uno causa dell’altro.

Per approfondire questo importante concetto puoi consultare:

- un video (<https://www.youtube.com/watch?v=Tg6e2Y3IEUk>) che spiega che anche se le vendite di gelati e le scottature solari sono correlate (cioè crescono insieme) non significa che l’una causa l’altra, ma piuttosto che un’altra causa (verosimilmente il calore del sole) le condiziona entrambe.

Qui <https://www.tylervigen.com/spurious-correlations> trovi esempi fantasiosi di forti correlazioni statistiche tra variabili che però non hanno nulla a che vedere l’uno con l’altro (ad esempio il consumo di latte con il tasso di divorzi in Colorado).

SCHEDA OPERATIVA – ANALIZZARE I DATI CON STRUMENTI ACCESSIBILI

A COSA SERVE

Questa scheda è pensata per chiunque possa utilizzare strumenti accessibili, come Excel o Google Sheets, e voglia svolgere prime analisi in modo ordinato, leggibile e metodologicamente prudente.

Non si tratta di una guida all'utilizzo di fogli di calcolo, per questo consigliamo guide esterne sulle specificità del loro utilizzo, ivi comprese guide sull'applicazione di formule tipicamente sfruttate nel lavoro con dati.

L'utilizzo di chatbot AI è altresì incoraggiato, anche se con occhio critico e sempre volto all'apprendimento e non alla sostituzione, per ottenere suggerimenti ad hoc sulle funzionalità di Excel o altri software, data l'elevata affidabilità che hanno raggiunto questi strumenti nel svolgere quanto verrà delineato. Pertanto, di seguito si illustreranno solo esempi di utilizzo e buone pratiche.

1. VERIFICARE CHE I DATI SIANO LEGGIBILI PRIMA DI ANALIZZARLI

Prima di interpretare i dati, occorre controllare che il file sia effettivamente utilizzabile. Anche un dataset ben costruito può contenere errori, valori incoerenti o categorie scritte in modo non uniforme.

Controlli essenziali

- verificare la presenza di celle vuote o incomplete;
- controllare se alcuni valori sono improbabili o fuori scala;
- individuare eventuali duplicati;
- verificare che le categorie siano scritte sempre nello stesso modo;
- controllare che ogni colonna contenga una sola informazione.

Esempio

In una colonna dedicata alla classe compaiono i valori 2A, 2 a, IIA. Prima di qualsiasi analisi queste etichette devono essere uniformate, altrimenti i conteggi per classe risulteranno falsati.

Strumenti utili in Excel

- **Filtro e Ordina** per esplorare rapidamente i valori;
- **Formattazione condizionale** per evidenziare duplicati o celle vuote;
- **CONTA.VUOTE** per contare i mancanti;
- **CONTA.SE** per verificare quante volte compare una categoria.
- **UNICI** per produrre una lista di valori unici in una categoria

Questa fase non serve ancora a spiegare i dati. Serve a verificare che ciò che si osserva sia plausibile e leggibile.

2. DESCRIVERE I DATI: OSSERVARE COSA EMERGE

Il primo livello di analisi consiste nel descrivere i dati. Descrivere significa riassumere in modo chiaro ciò che il dataset mostra, senza formulare ancora spiegazioni.

Domande utili

- quanti casi sono presenti nel dataset?
- quali sono i valori più frequenti?
- qual è il valore medio?
- i dati sono distribuiti in modo omogeneo oppure molto disperso?
- sono presenti casi molto distanti dagli altri?

Indicatori base

- numero di casi;
- minimo e massimo;
- media;
- mediana;
- frequenze per categoria.

Strumenti utili in Excel

- MEDIA
- MIN/MAX
- MEDIANA
- CONTA.NUMERI
- CONTA.SE

Esempio: Nel gruppo considerato, le assenze medie sono pari a 11. I valori osservati vanno da 0 a 38. La mediana è 9, quindi metà degli studenti presenta 9 o meno assenze.

Questa è una descrizione iniziale per capire le dimensioni del fenomeno.

3. CONFRONTARE GRUPPI PER INDIVIDUARE DIFFERENZE RILEVANTI

Una volta ottenuto un primo quadro descrittivo, si può passare al confronto tra gruppi. Nella scuola questo passaggio è spesso molto utile, perché consente di osservare differenze tra classi, sezioni, plessi o gruppi di studenti con caratteristiche diverse.

Confronti possibili

- tra classi;
- tra sezioni;
- tra indirizzi;
- tra studenti con diverse caratteristiche comuni;
- tra studenti con alta e bassa frequenza;
- tra partecipanti e non partecipanti a un progetto;
- tra periodi diversi;

Obiettivo

Verificare se emergono differenze che meritano attenzione, senza attribuire subito un significato causale.

Strumenti utili in Excel

- MEDIA.SE
- MEDIA.PIÙ.SE
- CONTA.PIÙ.SE
- Tabelle pivot

ESEMPIO OPERATIVO 1 – CONFRONTARE RISULTATI TRA CLASSI

Contesto: Si vogliono confrontare i voti di italiano tra classi.

Estratto Excel

	A	B	C	D	E	F	G
1	id_studente	anno classe	sezione	voto_italiano		classe	media_voti
2	XXXX	3	A	7		3A	6,5
3	XXXX	3	A	6		3B	4,5
4	XXXX	3	B	5		2A	8,5
5	XXXX	2	B	6		2B	6,5
6	XXXX	2	A	8			
7	XXXX	2	B	7			
8	XXXX	3	B	4			
9	XXXX	2	A	9			

Operazione: Calcolare la media dei voti per classe e sezione.

Come funziona: la formula **MEDIA.PIÙ.SE** calcola la media dei voti **solo per le righe che appartengono a una o più categorie specificate**. In questo caso, per la 3A, la formula utilizza nel calcolo della media le osservazioni che hanno valore "2" nella colonna classe e valore "A" nella colonna sezione.

ESEMPIO OPERATIVO 2 – CREARE GRUPPI A PARTIRE DAI DATI

Contesto: Le assenze sono espresse in numeri e risultano poco leggibili.

Estratto Excel

	A	B	C	D	E	F	G
1	fascia_assenze	n_assenze					
2	basse	<6					
3	medie	6-15					
4	alte	>15					
5							
6	id_studente	n_assenze	voto_mate	fascia_assenze		fascia_assenze	media_voti
7	XXXX	2	8	basse		basse	7,5
8	XXXX	6	6	medie		medie	6
9	XXXX	12	6	medie		alte	5,5
10	XXXX	22	5	alte			
11	XXXX	4	7	basse			
12	XXXX	17	6	alte			

Operazione: Suddividere le assenze in fasce (basse, medie, alte).

Come funziona: La funzione SE assegna una categoria a ciascun valore: ad esempio, <6 = “basse”, $6-15$ = “medie”, >15 = “alte”.

Questo permette di raggruppare i dati secondo i gruppi creati utilizzando, ad esempio, **MEDIA.SE** come visto nell'esempio precedente

Esempio di lettura: La media dei voti di matematica risulta più bassa tra gli studenti con assenze elevate rispetto agli altri gruppi. Questo dato segnala una differenza che merita approfondimento, ma non consente da solo di spiegare le ragioni di tale associazione.

USO CONSIGLIATO DELLE TABELLE PIVOT (AVANZATO)

Le tabelle pivot sono uno degli strumenti più utili per una prima analisi, perché permettono di compiere la maggior parte delle azioni menzionate automaticamente in una tabella che cambia forma in base alle informazioni che si vogliono evidenziare, ad esempio:

- riassumere velocemente un dataset;
- confrontare gruppi;
- produrre conteggi e medie;
- costruire basi semplici per grafici e discussione.

L'utilizzo di queste tabelle, tuttavia, non è intuitivo nell'immediato e potrebbe richiedere maggiore padronanza dei fogli di calcolo.

4. INDIVIDUARE ANOMALIE, ECCEZIONI E RISULTATI INATTESI

Analizzare i dati non significa solo cercare conferme. Al contrario, una buona analisi presta attenzione anche ai casi che non seguono l'andamento generale: valori estremi, risultati inattesi, situazioni che contraddicono l'ipotesi iniziale.

Perché è utile

Le eccezioni aiutano a leggere meglio il fenomeno e a evitare interpretazioni affrettate. Elementi da ricercare:

- errori nel dato;
- differenze di contesto;
- situazioni da approfondire con strumenti qualitativi;
- elementi utili per rivedere l'ipotesi di partenza.

Esempio: Se la maggior parte degli studenti con molte assenze presenta voti bassi, ma alcuni ottengono comunque buoni risultati, questi casi non vanno ignorati. Possono suggerire che il rapporto tra frequenza e rendimento sia più complesso di quanto sembri.

Strumenti utili in Excel

- ordinamento crescente o decrescente;
- filtri;
- MIN/MAX;
- formattazione condizionale;
- grafici.

I casi anomali non vanno eliminati automaticamente. Prima occorre capire se si tratta di un errore o di un caso reale che aggiunge complessità all'analisi.