

Statistica e Valutazione della Formazione

Slides 1

A.A. 2025-2026

CdL Magistrale in [SCIENZE PEDAGOGICHE](#)

Docente: ANNA LINA SARRA

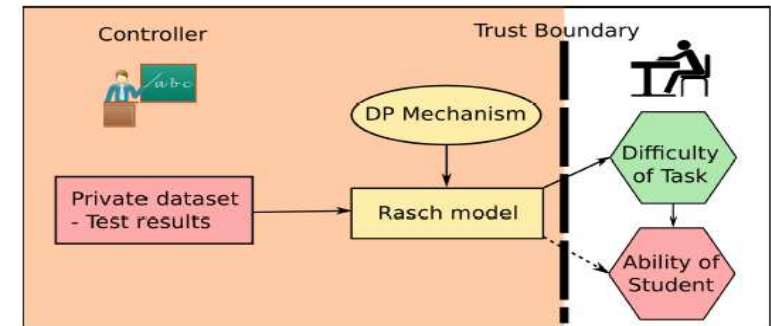
Obiettivo del corso

- Il corso si propone di fornire i principali metodi di analisi e valutazione dei sistemi formativi.
- Nella **prima parte del corso** verranno ripresi i concetti fondamentali della statistica descrittiva
- Nella **seconda parte** verranno introdotti i principali metodi statistici utilizzabili per la valutazione di sistema e degli apprendimenti in campo formativo.

INTRODUZIONE AL CORSO

Il programma prevede lo svolgimento dei seguenti moduli:

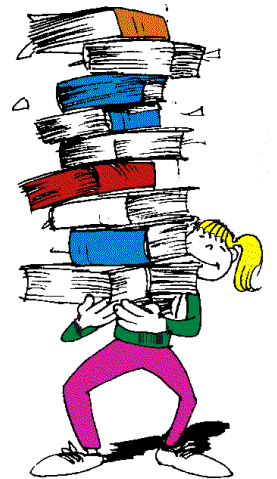
- **MODULO 1.** Statistica descrittiva
- **MODULO 2.** I metodi della valutazione quantitativa
(La teoria Classica dei Test;- L'Item Response Theory)
- **MODULO 3.** La progettazione del questionario



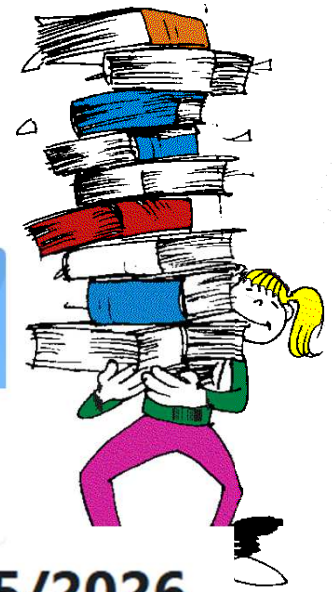
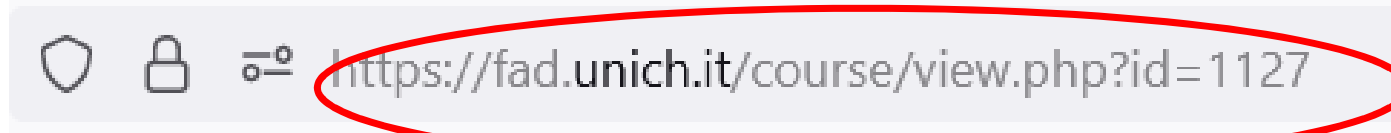
Materiale

Testi Consigliati

- Cicchitelli, G., D'Urso, P., & Minozzo, M. Statistica: principi e metodi, 4^a edizione (Ediz. MyLab, con aggiornamento online), Pearson, 2022.
- LA CENTRALITA' DELLA VALUTAZIONE NELLA VALORIZZAZIONE DELL'ISTRUZIONE DEL CAPITALE UMANO - C. CECCATELLI - FRANCOANGELI - ISBN 9788820458645
- Durante lo svolgimento delle lezioni sarà fornito materiale integrativo che potrà essere scaricato dalla piattaforma **fad.unich.it**



Materiale



[SVF25] - STATISTICA E VALUTAZIONE DELLA FORMAZIONE...
I Anno
[SVF25] - STATISTICA E VALUTAZIONE DELLA FORMAZIONE 2025/2026

[SVF25] - STATISTICA E VALUTAZIONE DELLA FORMAZIONE 2025/2026

Corso Impostazioni Partecipanti Valutazioni Report Altro ▾

▼ Introduzione

🗨️ Annunci

Modalità di erogazione

Per ogni argomento la lezione sarà suddivisa in

- **Aspetti teorici**
- **Esemplificazioni**
- **Per alcuni argomenti sono previste applicazioni in Excel.**



Esame

- La verifica finale consiste **in colloquio orale**.

Orario lezioni e contatti

LUNEDÌ	GIOVEDÌ	VENERDÌ
16—18 V DENTE LETTERE	14-16 AULA 8 LETT PD	11-13 AULA B GEOLOGIA (EX-RETTORATO)

- **Giorni ed orari di ricevimento studenti:** dopo le lezioni e per appuntamento da concordarsi via e-mail
- **E-mail:** annalina.sarra@unich.it

LA STATISTICA

La STATISTICA è....

La STATISTICA è la disciplina che elabora i principi e le metodologie che presiedono al processo di rilevazione e raccolta dei dati, alla rappresentazione sintetica e alla interpretazione dei dati stessi, e laddove ve ne siano le condizioni, alla generalizzazione delle evidenze osservate.

La definizione di STATISTICA

La statistica non è più la “scienza dello Stato”. Non esiste una definizione univocamente accettata, ma le varie proposte hanno spesso un ampio respiro.

Alcune definizioni

“La statistica riguarda ciò che si può imparare dai dati”. Anthony Davison.

“La statistica è essenzialmente lo studio dell'incertezza. Il ruolo degli statistici è di assistere i lavoratori di altri campi, o i propri clienti, i quali incontrano l'incertezza nel loro lavoro.” Dennis Lindley.

“La statistica è la tecnologia necessaria per estrarre significato dai dati e per gestire l'incertezza”. David Hand.

“La statistica è la scienza delle decisioni in condizioni di incertezza”. Domenico Piccolo

Verso la STATISTICA moderna

Il ruolo della statistica cambiò notevolmente a partire dalla fine del 1800, anche grazie allo sviluppo del calcolo delle probabilità e del gioco d'azzardo.

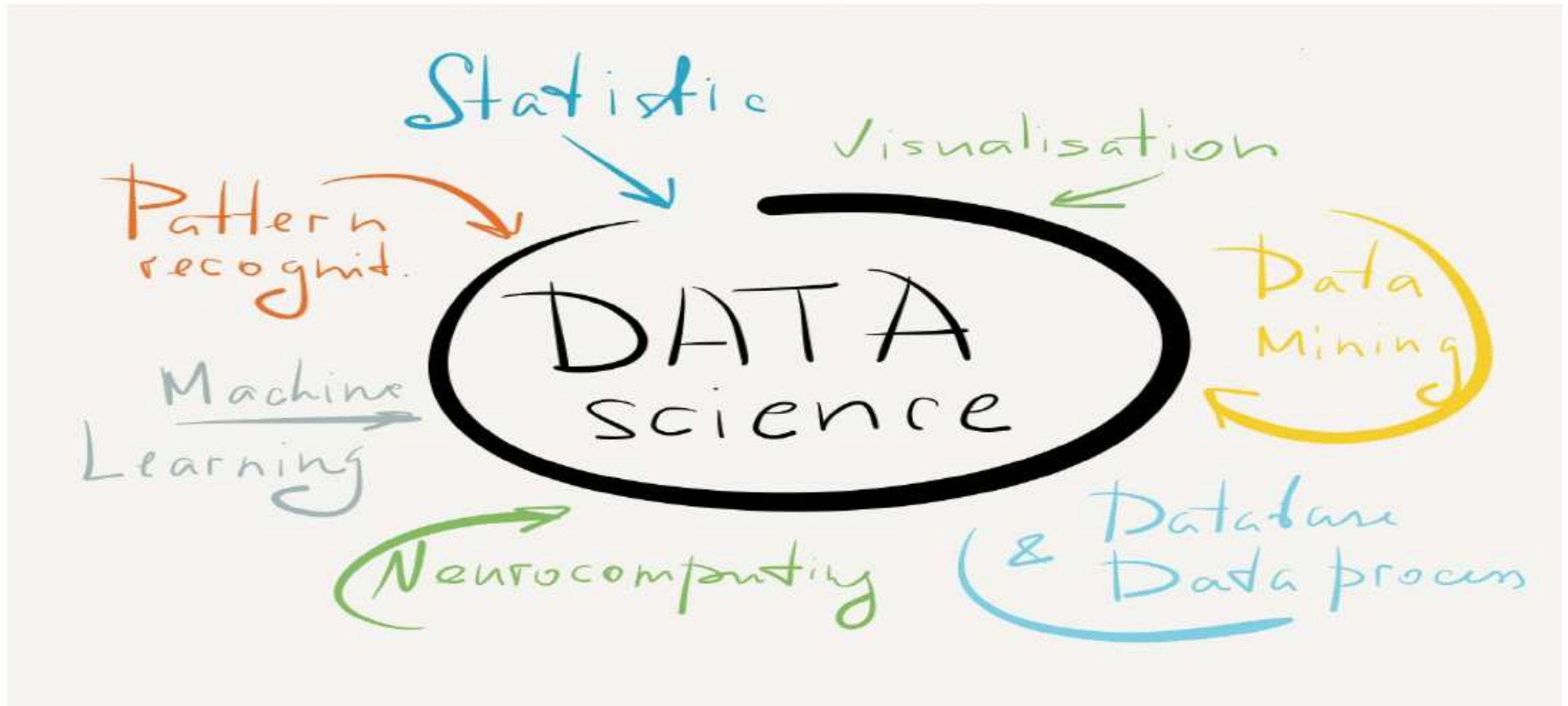
La Royal Statistical Society (RSS) venne fondata nel 1834, mentre l'American Statistical Association (ASA) nel 1839.

Il primo dipartimento di statistica al mondo venne fondato nel 1911 all'University College di Londra (UCL).

Nella prima metà del '900 avvenne la prima rivoluzione della statistica. Furono gettate le fondamenta della teoria della statistica moderna.

A partire dagli anni '70 e '80 del secolo scorso è cominciata una seconda rivoluzione, grazie all'immensa potenza di calcolo e disponibilità di dati derivanti dallo sviluppo informatico. Questa fase rivoluzionaria è ancora in atto. Nel contesto moderno, la statistica viene anche chiamata DATA SCIENCE.

DATA SCIENCE



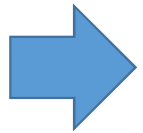
LA STATISTICA è UNA SCIENZA ANCILLARE

«L'aspetto migliore dell'essere uno statistico è che puoi giocare nel giardino di chiunque». (John Tukey).

La statistica è una scienza di supporto a tutte le altre discipline,
ad esempio:

Biologia, Medicina, Epidemiologia, Scienze sociali, Demografia, Psicologia Marketing, Finanza.

La statistica viene spesso chiamata una scienza ancillare, perché è presente trasversalmente in quasi tutte le discipline scientifiche.



Al tempo stesso, la statistica è una disciplina a sé stante, con le sue regole ed il suo linguaggio.

La percezione della statistica

- *“Ci sono tre tipologie di menzogne: le bugie, le maledette bugie e le statistiche”.* Pronunciata da **Mark Twain**.
- *“Come i sogni, le statistiche sono una forma di appagamento”.* **Jean Baudrillard**.

A volte purtroppo la statistica è accolta con sospetto.

La statistica e i dati, di per sé non mentono mai. Chi mente sono due categorie di statistici:

quelli **DISONESTI** e quelli **INCOMPETENTI**.

Sfortunatamente, anche quando un bravo statistico svolge correttamente il suo lavoro, questo potrebbe venire **PERCEPITO** come «falso» se mancano le competenze quantitative di base.

La statistica nella cultura popolare

La figura qui di fianco

illustra un esempio di statistico incompetente.



In risposta alle precedenti provocazioni

- Fortunatamente, non tutti sono dello stesso stesso parere. La statistica è in effetti **alla base del metodo scientifico**.
- *"E' semplice mentire con la statistica, ma è molto più semplice mentire senza di essa"*. **Frederick Mosteller**.
- *"La statistica è la grammatica della scienza"*. **Karl Pearson**.
- *"Coloro che non conoscono la statistica sono condannati a reinventarla"*. **Bradley Efron**.
- *"Il mestiere sexy dei prossimi 10 anni sarà lo statistico. Le persone pensano che stia scherzando, ma chi avrebbe immaginato che gli informatici sarebbero stati sexy negli anni '90?"* **Hal Varian**, chief economist di Google.

Il ruolo dello statistico

Ogni problema ha caratteristiche peculiari, con conoscenze pregresse diverse. In ogni problema statistico ci sono scelte da fare.

Non esistono ricette e procedure da applicare in automatico. Guardate ai contenuti dei corsi (di Statistica) solo come ad esempi di quello che si può fare.

Del resto, perché dovrebbero esistere delle lauree in Statistica se gli statistici fossero sostituibili con delle macchine?

Questo corso mostra idee, tecniche e strategie che si sono rivelate utili nel contesto educativo.

Ma i problemi che voi affronterete in futuro non saranno gli stessi.

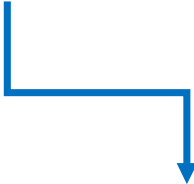
La statistica a chi serve e perchè

I. In generale abbiamo a che fare con le statistiche:

il PIL, il tasso di disoccupazione, l'indice dei prezzi e il tasso di inflazione, l'indice della produzione industriale, il tasso di dispersione scolastica

UTILIZZATA IN MODO
PROFESSIONALE O
STRUMENTALE

LA STATISTICA DEVE
ESSERE CONOSCIUTA



Dietro quei dati c'è la statistica:

metodi di campionamento, costruzione dei questionari, tecniche di somministrazione, tecniche di analisi dei dati semplici e complesse, valutazioni probabilistiche

Alcuni campi di applicazione

- STUDI DI POPOLAZIONE-DEMOGRAFIA
- MODELLI ECONOMETRICI
- MODELLI DI FINANZA
- EPIDEMIOLOGIA
- MODELLI AMBIENTALI (STUDI QUALITA' DELL'ARIA)
- SOCIOLOGIA (INDAGINI SU ATTEGGIAMENTI, MOTIVAZIONI ETC...)
- PEDAGOGIA (VALUTAZIONE DELL'APPRENDIMENTO)
- LEGAL STATISTICS (ANALISI DEL DNA DA TRACCE LASCIATE SUI LUOGHI DI DELITTI)
- ETC.....

Perche' la Statistica ?

Pensiamo statisticamente=> fin da bambini classifichiamo, sistemiamo, decidiamo

Gestione dell'incertezza => i fenomeni fisici, naturalistici, sociali, economici, ... sono aleatori. Occorre orientarsi e prendere decisioni.

Transizione delle discipline sostanziali da qualitative a quantitative, da deduttive a induttive=> salto di qualità, avanzamenti

Perche' la statistica

- PER CAPIRE
- PER VALUTARE
- PER SCEGLIERE/DECIDERE
- PER PREVEDERE/OPERARE
- PER SODDISFARE CURIOSITA'
- PER NON FARSI IMBROGLIARE

La statistica è intorno a noi

Punteggi e recensioni

4,5 ●●●●● 231 recensioni

Recensioni (231)

Controlliamo le recensioni ⓘ

Valutazione



Tipo di viaggiatore

- ☐ Famiglie
- ☐ In coppia
- ☐ Da solo
- ☐ Affari
- ☐ Amici

Periodo dell'anno

- ☐ Mar-Mag
- ☐ Giu-Ago
- ☐ Set-Nov
- ☐ Dic-Feb

Punteggi e recensioni

3,0 ●●●○○ 813 recensioni

Recensioni (813)

Controlliamo le recensioni ⓘ

Valutazione



Tipo di viaggiatore

- ☐ Famiglie
- ☐ In coppia
- ☐ Da solo
- ☐ Affari
- ☐ Amici

Periodo dell'anno

- ☐ Mar-Mag
- ☐ Giu-Ago
- ☐ Set-Nov
- ☐ Dic-Feb

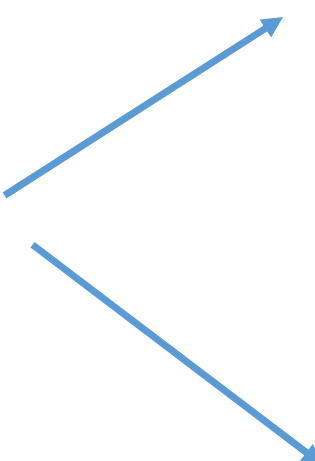
Poniamoci alcune domande

- ✓ Qual è la fonte delle informazioni?
- ✓ Come vengono raccolti i dati?
- ✓ Sulla base di cosa faccio la mia scelta?
- ✓ Come posso leggere i dati?

Come analizzare un fenomeno?

L'attività scientifica in generale si basa sull'osservazione empirica della realtà.

Distinguiamo tra:



-Ricerca quantitativa (uso di tecniche più strutturate e basata su un numero più ampio di osservazioni)

-Ricerca qualitativa (più orientata alla comprensione di un fenomeno che alla quantificazione)

Le fasi di una ricerca

- ✓ Progettazione o disegno della ricerca
- ✓ Raccolta/rilevazione dei dati
- ✓ Analisi dei dati
- ✓ Diffusione dei risultati

Progettazione o disegno della ricerca

- Definire operativamente i concetti della ricerca;
- Definire gli oggetti della ricerca (le unità di rilevazione);
- Stabilire se condurre una rilevazione su tutte le unità o su una parte (rilevazione censuaria/campionaria);
- Stabilire le modalità di passaggio dal piano teorico a quello empirico;
- Costruire lo strumento di rilevazione (ad es. un questionario);
- Prefigurare la fase di analisi dei dati.

Raccolta dei dati

È la cosiddetta fase sul campo di una ricerca.

Si raccolgono i dati presso le unità di rilevazione. Il risultato è una matrice dei dati, ovvero una tabella dove in riga troviamo le risposte dei singoli e in colonna troviamo le risposte alle singole variabili.

Nell'incrocio tra un riga e una colonna troviamo un dato.

Esempio di matrice dei dati (unità x variabili)

ID	TIPOLOGIA DI INSEGNANTE	ETA'	ANNI DI SERVIZIO	ORDINE SCUOLA IN CUI INSEGNA	GENERE	La disciplina che ho studiato è stata decisiva nella mia scelta della professione di insegnante.	Possiedo una solida conoscenza della mia disciplina, per il mio lavoro di insegnante.
1	tutor di un neoassunto	41	11	Primaria	femminile	Molto d'accordo	D'accordo
2	tutor di un neoassunto	51	23	Secondaria di secondo grado	maschile	Molto d'accordo	Molto d'accordo
3	tutor di un neoassunto	34	5	Secondaria di primo grado	maschile	Molto d'accordo	D'accordo
4	tutor di un neoassunto	48	8	Secondaria di secondo grado	maschile	D'accordo	D'accordo
5	tutor di un neoassunto	45	10	Secondaria di secondo grado	maschile	Molto d'accordo	D'accordo
6	tutor di un neoassunto	31	6	Secondaria di secondo grado	maschile	Molto d'accordo	Molto d'accordo
7	tutor di un neoassunto	56	30	Secondaria di primo grado	femminile	Molto d'accordo	D'accordo
8	tutor di un neoassunto	46	19	Secondaria di secondo grado	femminile	Molto d'accordo	Molto d'accordo
9	tutor di un neoassunto	47	16	Secondaria di primo grado	femminile	Molto d'accordo	Molto d'accordo
10	tutor di un neoassunto	40	11	Primaria	femminile	D'accordo	Molto d'accordo
11	tutor di un neoassunto	61	40	Secondaria di primo grado	femminile	Abbastanza d'accordo	D'accordo
12	tutor di un neoassunto	55	26	Secondaria di secondo grado	maschile	Abbastanza d'accordo	D'accordo
13	tutor di un neoassunto	64	39	Secondaria di secondo grado	femminile	Molto d'accordo	Molto d'accordo
14	tutor di un neoassunto	53	15	Primaria	femminile	D'accordo	D'accordo

Raccolta dei dati

Il dato si definisce come:

«l'esito di un processo di rilevazione delle informazioni alla luce dell'apparato metodico e concettuale di una certa disciplina».

Abbiamo visto che è il contenuto di una singola cella di una matrice dei dati.

Per poter parlare di dato dev'esserci una definizione operativa, ovvero quell'insieme di regole per poter trasformare i possibili stati di una variabile in un numero o descrizione da inserire nella matrice.

Tre livelli dei dati

MICRODATI:

costituiscono l'unità elementare di informazione, ovvero un dato sulla singola unità statistica (ad es. il singolo voto ad un esame).

MACRODATI:

sono unità complesse di informazione, costituite dall'elaborazione e dalla sintesi dei microdati (ad es. la media dei voti di una sessione di esame).

METADATI:

sono i «dati sui dati». Forniscono informazioni dettagliate sui dati e sul loro processo di produzione per consentire una corretta lettura e interpretazione (ad. es. periodo e modalità di raccolta, definizioni operative, ecc.).

In pratica...

MICRODATI: Lorenzo pesa 70 kg ed è alto 1,76 m.

MACRODATI: il 44,4% delle persone di 18 anni e più in Italia è sovrappeso o obesa.

METADATI: Proporzione (standardizzata con la popolazione europea al 2013) di persone di 18 anni e più in sovrappeso o obese sul totale delle persone di 18 anni e più.

L'indicatore fa riferimento alla classificazione dell'Organizzazione mondiale della sanità (OMS) dell'Indice di Massa corporea (Imc: rapporto tra il peso, in Kg, e il quadrato dell'altezza, in metri).

Fonte: Istat, Indagine Aspetti della vita quotidiana. Anno 2021.

Le fonti dei dati

I dati possiamo raccoglierli direttamente (dati primari), in altri casi possiamo utilizzare quelli raccolti da altri (dati secondari).

La principale fonte di dati secondari in Italia è l'Istat, l'Istituto Nazionale di Statistica, cui fa capo il Sistan, il sistema statistico nazionale.

Dal sito di Istat è possibile scaricare una grande quantità di dati, sotto forma di microdati e macrodati. Sono disponibili anche i metadati, per un uso più corretto e consapevole dei dati.

Esistono svariate fonti di dati, come ad es. gli **istituti privati di ricerca, le università**, ecc., ma a volte facciamo ricorso a dati prodotti non per finalità statistiche, ma per altri motivi (ad es. dati amministrativi).

Infine, negli ultimi anni si parla sempre più spesso di **BIG DATA**, cioè grandi quantità di dati generate automaticamente da attività digitali, come l'uso dei social media, dei telefoni cellulari o delle piattaforme online.

I BIG DATA

Si definisce col termine big data una raccolta di dati così estesa in termini di volume, velocità e varietà da richiedere tecnologie e metodi analitici specifici per l'estrazione di valore.



I recenti progressi tecnologici hanno ridotto esponenzialmente il costo di archiviazione e di elaborazione dei dati, rendendo più facile e meno costoso archiviare più dati rispetto al passato.

Con un volume crescente di big data, diventato sempre più economico e accessibile, si possono conoscere i fenomeni sociali, economici, ecc. in modo più approfondito e prendere decisioni e fare valutazioni più accurate e precise.

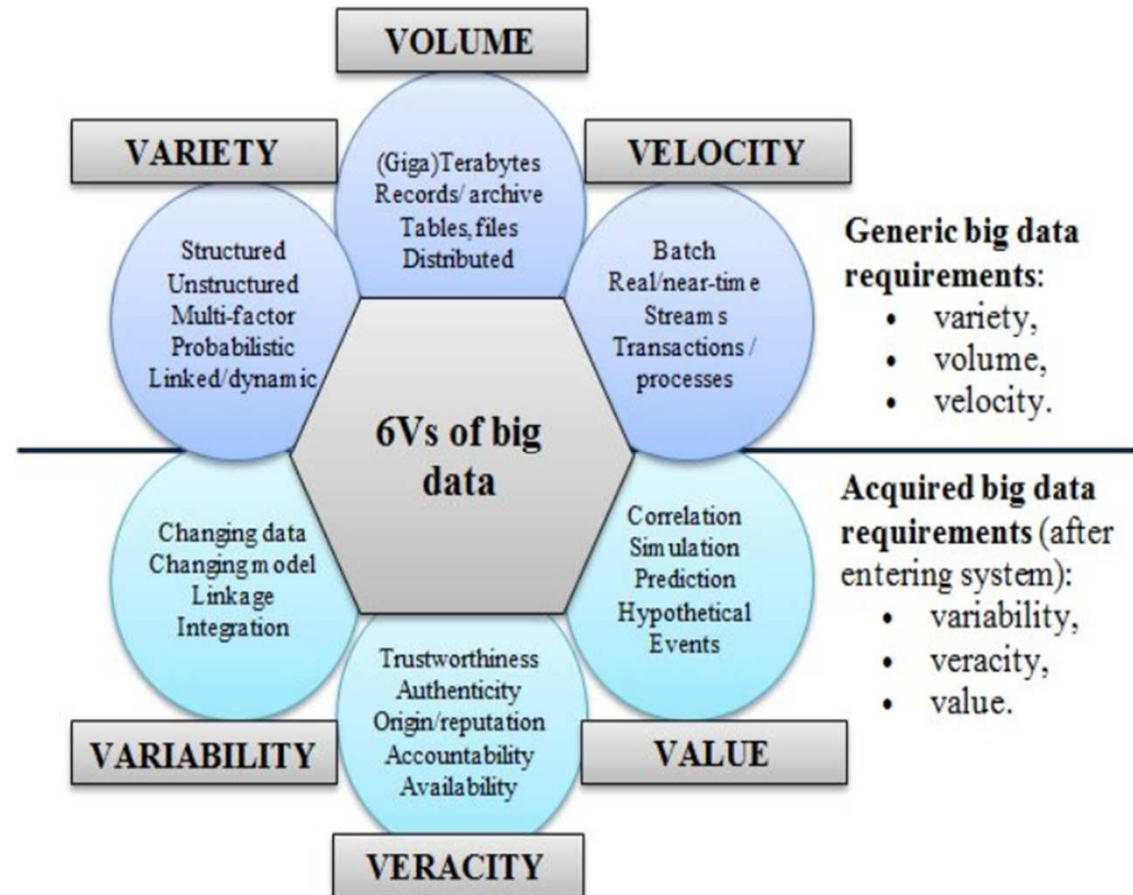
I BIG DATA e le 6 V

Dati provenienti principalmente da **Social Media** e/o dispositivi tecnologici.



Paradigma di ricerca diverso rispetto a ciò che abbiamo detto.

Nella ricerca educativa classica si raccolgono dati con strumenti controllati (questionari, osservazioni, interviste). Con i Big Data invece i dati esistono già, sono prodotti continuamente e spesso senza che nessuno li abbia progettati a scopo di ricerca. Questo cambia profondamente il modo di fare domande, di interpretare i risultati e di ragionare sull'etica della ricerca.



I BIG DATA: esempio una piattaforma di e-learning scolastica

Immaginate che una scuola adotti una piattaforma digitale per tutti i suoi studenti.

Ogni giorno, automaticamente, la piattaforma registra:



Chi si collega
e quando



Per quanto tempo
resta connesso



Quali materiali
ha aperto



Quante volte
rilegge un testo



A che ora
consegna i compiti



Quanti errori
fa in un quiz

Ecco come entrano in gioco le 6 V dei Big Data.

Volume--Quanti dati?

? Definizione

Volume è la quantità enorme di dati generati continuamente. È il criterio principale per parlare di Big Data: se la quantità è tale da non poter essere gestita con strumenti tradizionali, siamo di fronte a un fenomeno di Big Data.

? Nella nostra scuola con 500 studenti:

Con 500 studenti che lavorano ogni giorno sulla piattaforma, in un solo mese si accumulano milioni di eventi registrati: accessi, click, risposte, consegne, visualizzazioni. Nessun insegnante potrebbe leggerli uno per uno. Serve un sistema automatico per gestirli.

? Il problema del Volume non è solo tecnico. Pedagogicamente significa: non possiamo più leggere ogni traccia lasciata da ogni studente. Chi decide cosa tenere e cosa scartare?

Velocity—Quanto velocemente arrivano?

Definizione

Velocity è la velocità con cui i dati vengono generati e devono essere elaborati. Non si tratta solo di dati già raccolti: i dati arrivano in tempo reale e spesso devono essere analizzati immediatamente per essere utili.

Nella nostra scuola con 500 studenti:

I dati non aspettano: arrivano mentre gli studenti lavorano, minuto per minuto. Se voglio capire se uno studente sta avendo difficoltà oggi, non posso aspettare il registro di fine mese. La Velocity è la sfida di lavorare con dati che scorrono continuamente.

? *Un sistema che rileva in tempo reale che uno studente ha sbagliato 10 volte di fila sullo stesso tipo di esercizio potrebbe intervenire subito. Ma chi ha il diritto di intervenire, e come?*

Variety—in che forma arrivano?

Definizione

Variety si riferisce alla diversità dei tipi di dati: strutturati (numeri, tabelle), semi-strutturati (log di sistema) e non strutturati (testi, audio, video). La sfida è integrare fonti di natura completamente diversa in un'unica analisi.

Nella nostra scuola con 500 studenti:

Sulla piattaforma convivono dati strutturati (voto di un quiz, numero di accessi), semi-strutturati (log di sistema) e non strutturati (testi nei forum, commenti vocali, video caricati dagli studenti). Non si tratta di numeri puliti in una tabella, ma di materiali di natura completamente diversa.

? *Un testo scritto da uno studente in un forum di discussione vale quanto un voto numerico? Come si confrontano queste informazioni qualitative con i dati quantitativi?*

Le prime 3 V:

descrivono i dati così come arrivano.

Sono i problemi che esistono ancora prima di iniziare ad analizzare qualcosa.

✓ Volume

✓ Velocity

✓ Variety

Le prossime 3 V:

emergono nel momento in cui provate a usare quei dati per capire qualcosa.

Sono i problemi dell'interpretazione, dell'affidabilità e dell'utilità.

Variability—I dati significano sempre la stessa cosa?

Definizione

Variability si riferisce ai cambiamenti nel significato, nel formato o nella velocità dei dati nel tempo e nel contesto. Lo stesso dato può avere interpretazioni completamente diverse a seconda di quando, dove e come viene raccolto.

Nella nostra scuola con 500 studenti:

Uno studente che si collega alle 23:00 sta studiando con impegno o sta copiando all'ultimo momento? Un tempo di lettura di 30 secondi su una pagina significa che l'ha letta velocemente perché è brava, o che non l'ha letta affatto? Lo stesso dato cambia significato a seconda del contesto.

? *Come pedagogisti, dovete sempre chiedervi: il contesto in cui interpreto questo dato è lo stesso in cui è stato prodotto? Un dato decontestualizzato può portare a conclusioni sbagliate.*

Veracity—Di quei dati mi possono fidare?

Definizione

Veracity si riferisce all'affidabilità e alla qualità dei dati. Dati imprecisi, incompleti o non comparabili portano a conclusioni errate, anche se l'analisi è tecnicamente corretta. È il problema della 'spazzatura in entrata, spazzatura in uscita'.

Nella nostra scuola con 500 studenti:

Magari uno studente lascia la piattaforma aperta e va a fare altro: il sistema registra 45 minuti di 'studio' che in realtà non sono avvenuti. Oppure i dati di due classi non sono comparabili perché i docenti hanno configurato la piattaforma in modo diverso. Questi dati riflettono davvero la realtà?

? *In una ricerca educativa, dati non affidabili possono portare a politiche scolastiche sbagliate. Chi verifica la qualità dei dati prima che vengano usati per prendere decisioni sugli studenti?*

Value--Tutto questo serve a qualcosa?

Definizione

Value è la V più importante: milioni di dati sono inutili se non producono conoscenza utile. I dati hanno valore solo quando vengono trasformati in informazioni che permettono di prendere decisioni migliori o di capire meglio un fenomeno.

Nella nostra scuola con 500 studenti:

Avere milioni di dati non significa automaticamente capire come stanno apprendendo gli studenti. Il valore emerge solo quando si fanno le domande giuste: quali studenti rischiano di abbandonare? Quali attività producono apprendimento migliore? In quali momenti la concentrazione è più alta?

? *Chi decide quale domanda fare determina quale valore viene estratto. E chi beneficia di quel valore: lo studente, il docente, la scuola, o chi ha costruito la piattaforma?*

Cosa ci dice la piattaforma scolastica?

Volume

Milioni di eventi/mese con 500 studenti

→ Serve gestione automatica

Velocity

Dati in tempo reale, minuto per minuto

→ Serve analisi immediata

Variety

Voti, log, testi, audio, video insieme

→ Serve integrazione di formati diversi

Variability

23:00 = studio serio o copia? Dipende dal contesto

→ Serve interpretazione contestuale

Veracity

Piattaforma aperta ≠ studio reale

→ Serve verifica della qualità

Value

Dati utili solo con le domande giuste

→ Serve un obiettivo pedagogico chiaro

Paradigma di ricerca diverso

Nella ricerca tradizionale:

si progetta uno strumento → si raccolgono dati in modo controllato → si analizzano.

Con i Big Data:

i dati esistono già, sono enormi, veloci, disordinati e non sempre affidabili.

- ▶ Questi dati su chi ci dicono qualcosa?
- ▶ A chi appartengono? Chi decide come usarli?
- ▶ Servono davvero agli studenti, o solo all'istituzione?

Raccolta dei dati

Certi fenomeni si possono misurare direttamente, come ad esempio l'età; altri necessitano di una fase detta **operazionalizzazione**, ovvero di essere trasformati in una o più domande, perché altrimenti il concetto o il fattore in sé non è misurabile.

❓ Pensiamo ad esempio ad alcuni concetti come povertà, benessere, salute, ...

❓ Esempi nel settore educativo

Concetto astratto	Possibili Indicatori
Benessere degli studenti	Soddisfazione scolastica, stress percepito
Motivazione allo studio	Tempo dedicato allo studio, partecipazione alle attività
Qualità dell'insegnamento	Valutazioni degli studenti, risultati degli apprendimenti
Competenze degli studenti	Punteggi nei test, risultati nelle prove standardizzate

L'analisi dei dati

Definizione

Per analisi dei dati si intende «la scelta delle tecniche statistiche più appropriate in base alle variabili a disposizione del ricercatore e ai suoi obiettivi conoscitivi».

Distinguiamo due grandi obiettivi conoscitivi

Obiettivo 1 — Descrittivo

Descrivere, esplorare e sintetizzare i dati raccolti.

L'obiettivo è rappresentare in modo chiaro ciò che si osserva nel campione o nella popolazione analizzata.

Obiettivo 2 — Inferenziale

Inferire i risultati ottenuti su un numero ristretto di unità all'intera popolazione.

L'obiettivo è generalizzare le conclusioni oltre il campione osservato.

Statistica descrittiva ed inferenziale

La statistica fa riferimento a due aree distinte ma collegate della conoscenza:

1 • Statistica descrittiva

Analisi di fenomeni collettivi a fini descrittivi.

 **Obiettivo: descrivere e sintetizzare**

Valori, coefficienti, tabelle, grafici per rappresentare i dati in forma compatta e leggibile.

2 • Statistica inferenziale

Analisi inferenziale per studiare e quantificare l'incertezza.

 **Obiettivo: generalizzare alla popolazione**

Si basa sulla teoria delle probabilità per passare dall'osservazione di un campione a tutta la popolazione.

Statistica descrittiva vs statistica inferenziale

Statistica descrittiva

- ▶ **Describe** e riassume la realtà.
 - Creazione di indici: media aritmetica, mediana, moda, varianza, ...
 - Visualizzazione: istogrammi, boxplot, ...
- ▶ **Quantifica** tramite strumenti algebrici, geometrici e grafici.
- ▶ **Esplora** i dati.

✦ Studia il campione o la popolazione per descriverne le caratteristiche in modo sintetico e visivo.

Statistica inferenziale

- ▶ **Quantifica** l'incertezza tramite il calcolo delle probabilità.
- ▶ **Stima** le quantità ignote di un modello statistico.
- ▶ **Prevede** fenomeni futuri fornendo margini di incertezza.

✦ Parte dai dati osservati su un campione per trarre conclusioni sull'intera popolazione, quantificando l'incertezza.

Fenomeni collettivi

Definizione

Il principale obiettivo della statistica descrittiva è la conoscenza quantitativa dei **fenomeni collettivi**: quei fenomeni che la nostra mente non può conoscere con una sola osservazione, ma che invece apprende tramite la **sintesi delle osservazioni di fenomeni più semplici**.

Esempi di fenomeni collettivi

- ▶ Disoccupazione dei laureati tra 24 e 35 anni in Italia
- ▶ Recidiva nei reati di violenza sessuale in Italia
- ▶ Evasione fiscale nelle piccole e medie imprese in Abruzzo
- ▶ La dispersione scolastica

In ambito educativo

È naturale che la statistica sia impiegata in **ambito educativo**, contraddistinto da una **molteplicità di dati** che richiedono strumenti di sintesi e analisi.

L'analisi dei dati

A seconda del numero di variabili prese in considerazione si parla di:

Analisi univariata se analizziamo le variabili di un fenomeno **una alla volta**.

Analisi bivariata se l'obiettivo è quello di studiare il legame tra **due** variabili.

Analisi multivariata se vogliamo studiare il legame tra **tre o più variabili**.

Attenzione! Il ruolo del
ricercatore è fondamentale:



«I numeri, a torturarli a lungo, confessano qualsiasi cosa»

Ronald Harry Coase (Willesden, 29 dicembre 1910 – Chicago, 3 settembre 2013), economista inglese naturalizzato statunitense, vincitore del premio Nobel per l'economia nel 1991.

La diffusione dei risultati

Restituzione dei principali risultati della ricerca e le conclusioni a cui si è giunti. In considerazione del pubblico a cui ci si rivolgerà cambieranno le modalità di restituzione: verranno prodotti grafici, tabelle e diagrammi con le relative interpretazioni.



LA STATISTICA IN AMBITO EDUCATIVO

La statistica in ambito educativo

- È interessante considerare in ambito educativo il ruolo della **statistica nei processi decisionali**.

- Nell'esperienza quotidiana, nella pratica e nella progettazione di **attività formative**, molteplici situazioni pongono dinnanzi alla necessità di formulare decisioni:
 - alcune sono orientate alla luce di un semplice ragionamento logico;
 - altre invece postulano informazioni precise, interpretate in maniera corretta

La statistica in ambito educativo

Gli strumenti e i metodi della **Statistica generale** (descrittiva e inferenziale), e dell'analisi **statistica multivariata** (in presenza di una pluralità di dati provenienti da una stessa unità di rilevazione) permettono di affrontare razionalmente ed efficacemente una **varietà di esperienze in ambito educativo**.

1. Ricerche pedagogiche fanno riferimento a dati statistici, punteggi a prove di esame, a risultati di indagine
2. Gli insegnanti valutano e sono valutati, misurano e sono oggetto di misurazione;
3. Nelle classi si impiegano dati espressi in forma qualitativa o quantitativa
4. Gli insegnanti applicano prove standardizzate di conoscenza o per la verifica di altre abilità e competenze

La statistica in ambito educativo

- In ambito educativo la statistica costituisce un valido strumento operativo che consente di trasformare in informazioni utili i dati forniti da fenomeni osservati allo scopo di

1. INTERPRETARE I FENOMENI
2. ORIENTARE LA FORMAZIONE DI DECISIONI

La decisione come processo

Ogni **decisione**, sia essa individuale o collettiva, personale o istituzionale, quale espressione di un impegno ed intervenire nella realtà, é il **prodotto ultimo di un insieme di attività di analisi**, elaborazione e confronto che, per la sua articolazione ed organicità, assume la configurazione di un **vero e proprio processo**.

Ciascun processo decisionale è riconducibile ad una sequenza di tre fasi :

- 1) Identificazione del problema
- 2) Formulazione di soluzioni
- 3) Adozione di una decisione

Identificazione del problema

- In presenza di un problema reale, che richiede una decisione, l'educatore deve formulare chiaramente in cosa esso consiste
- **È fondamentale tradurre i quesiti ambito educativo in quesiti statistici.**

• Inizia così la **prima fase del ragionamento statistico** e della ricerca statistica che è, in genere, detta di comprensione della problematica oggetto di studio, nella quale è importante capire bene quali sono i quesiti ai quali si deve rispondere in termini statistici ed acquisire sufficiente conoscenza del contesto educativo nel quale i quesiti sono formulati.

Formulazione di soluzioni

- Per poter implementare la seconda fase del ragionamento statistico, cioè la formulazione di soluzioni, lo statistico deve conoscere bene i **due ingredienti** per poter effettuare le analisi:



le informazioni statistiche

(fonti, caratteristiche e qualità dei dati disponibili e/o di quelli da rilevare con apposite indagini)

i metodi statistici

(sia descrittivi che inferenziali) che potrebbero adeguatamente essere utilizzati

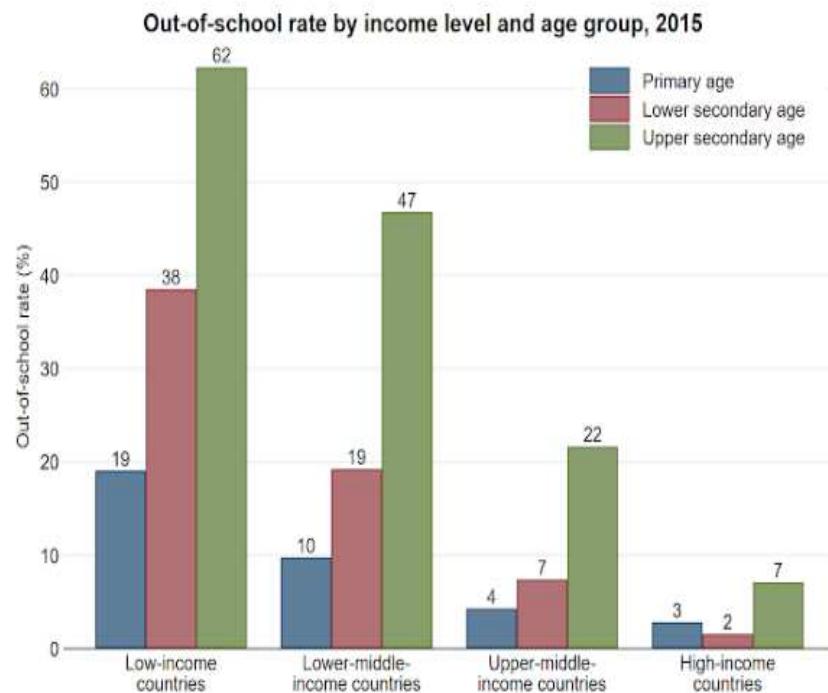
Adozione soluzione

- La terza e conclusiva fase del ragionamento statistico consiste **nell'interpretazione e nella valutazione della validità dei risultati** (outcome) dell'analisi in relazione all'iniziale problema ed ai quesiti statistici , e si conclude con la messa a disposizione dell'educatore delle risposte a tali quesiti.



Nel caso in cui il l'educatore non ritenga adeguate le risposte ai propri quesiti è necessario condurre nuovamente le analisi statistiche in relazione alle nuove richieste. Si procederà in questo modo fino a quando l'educatore ottiene risposte adeguate ai propri quesiti.

Esempi di utilizzo della statistica in education (1): dati di tipo macro



Friedrich Huebler, huebler.blogspot.com, July 2017

Where Do People Fail To Attain A High School Qualification?

% of 25-34 year-olds with an education level below upper secondary in selected countries*

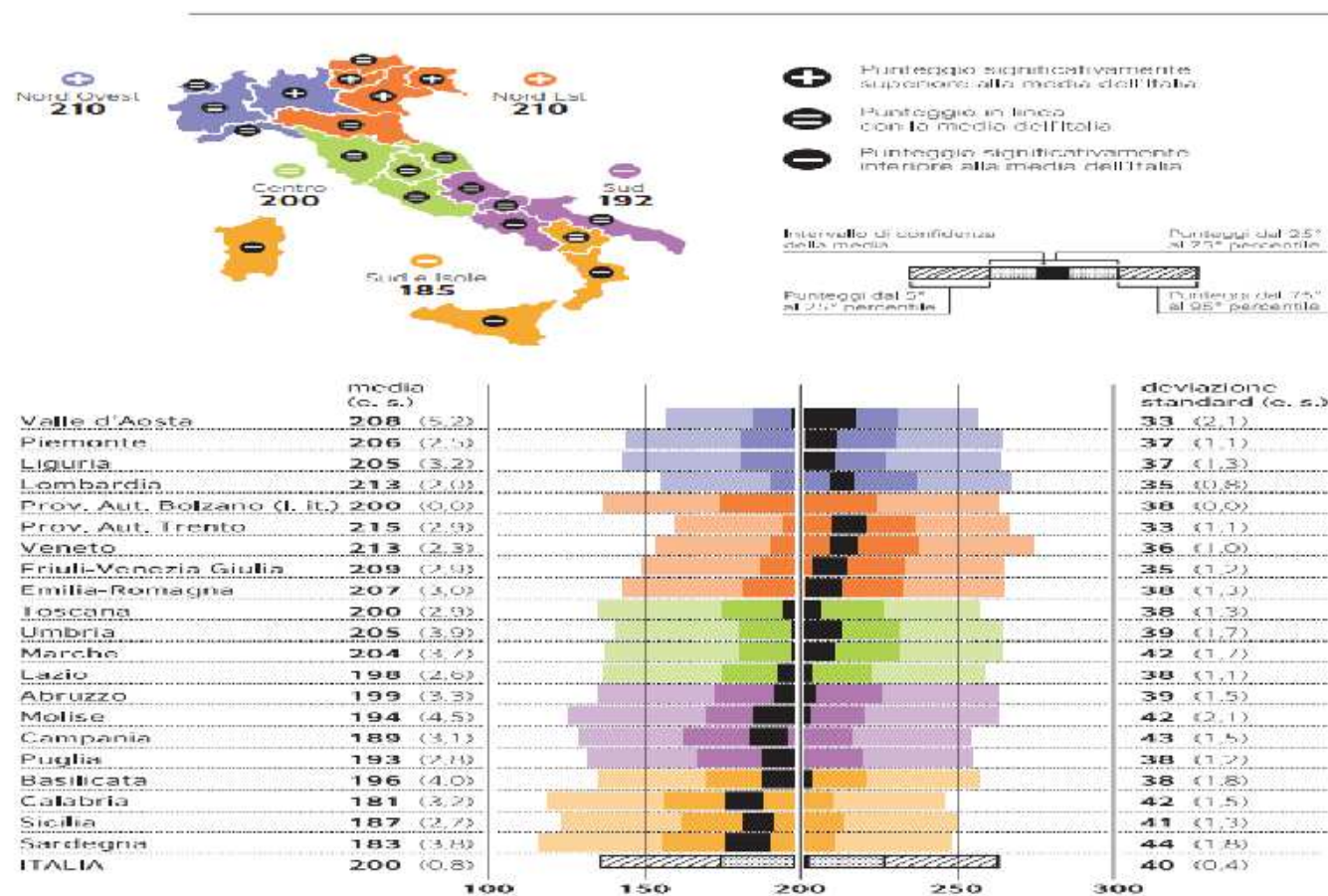


*2014 or latest year - upper secondary education corresponds to the final stage of secondary education in most OECD countries.
Source: OECD

CC BY-SA
@StatistaCharts

statista

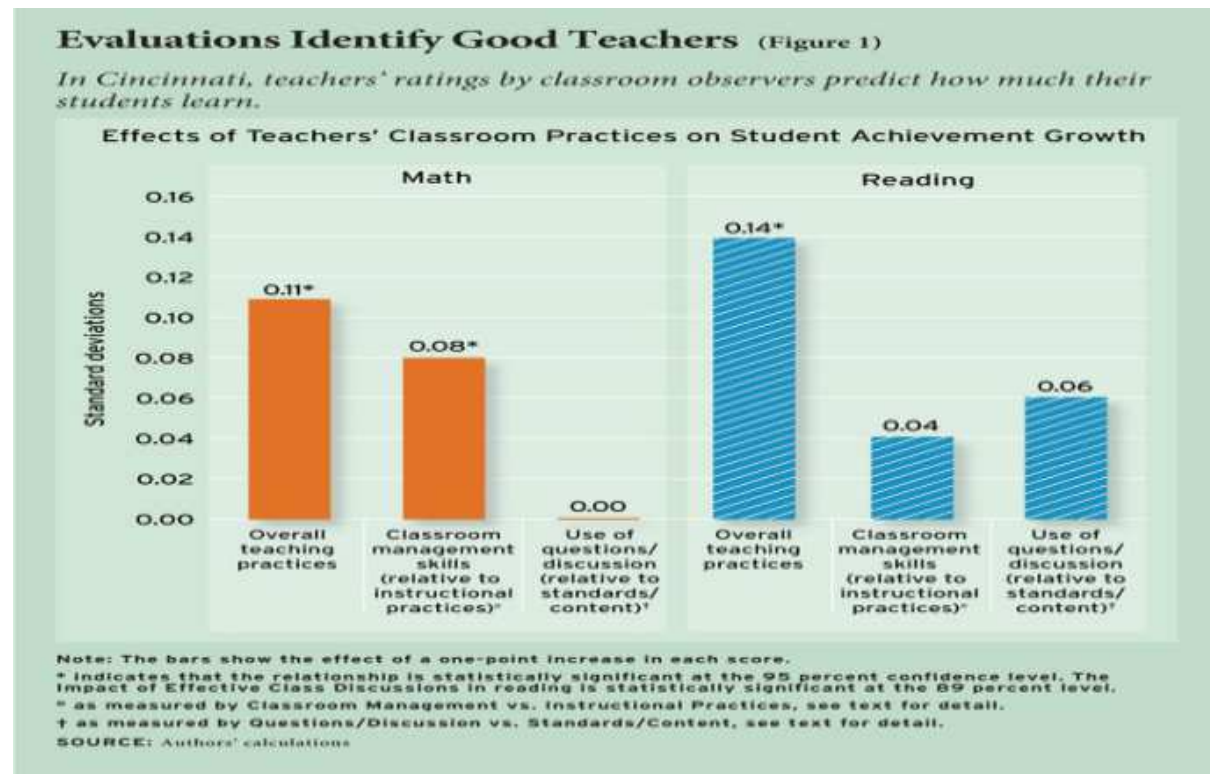
Esempi di utilizzo della statistica in education (2): dati di profitto



Dati INVALSI sono ampiamente analizzati per valutare eccellenze e criticità nel sistema formativo scolastico.

Esempi di utilizzo della statistica in education (3): modelli

Modelli statistici aiutano (ad esempio) a comprendere efficacia di diverse metodologie di insegnamento:



Funzione principale della statistica

La funzione principale della Statistica è quella di rendere comprensibile ciò che nella massa delle informazioni appare indifferenziato.

Quando si raccolgono le informazioni di un certo fenomeno, si ha a che fare con una mole notevole di dati grezzi.

Il primo problema da affrontare, quindi, è quello di sintetizzare la massa di dati grezzi in indicatori particolarmente informativi, utilizzando metodiche numeriche o grafiche, che siano in grado di descrivere le informazioni raccolte senza alternarne il senso complessivo.

Ma serve la statistica?

In una scuola secondaria di primo grado, gli insegnanti hanno realizzato una indagine sui viaggi e le vacanze estive degli alunni.

Hanno chiesto a ciascun ragazzo di esprimersi tra sei tipologie di possibili scelte: **nessun viaggio, mare, montagna, campagna, città d'arte, e altro**

Ma quanti sono?
Qual è stata la meta preferita?...
Siamo in grado di rispondere in pochi secondi?

mare	città d'arte	mare	mare	mare
mare	altro	città d'arte	montagna	nessun viaggio
nessun viaggio	mare	altro	altro	nessun viaggio
montagna	mare	mare	montagna	mare
mare	nessun viaggio	nessun viaggio	mare	montagna
nessun viaggio	nessun viaggio	nessun viaggio	mare	campagna
campagna	montagna	nessun viaggio	altro	campagna
città d'arte	nessun viaggio	altro	città d'arte	montagna
nessun viaggio	mare	mare	montagna	altro
mare	mare	mare	nessun viaggio	città d'arte

Ma se mettiamo tutto in una tabella.....



I dati che prima erano in una successione ora sono ben riassunti da una tabella!

Gli studenti in totale sono 50.

Mentre la meta preferita è stata il mare!

Tipo di vacanza	Totale
mare	17
nessun viaggio	12
montagna	7
altro	6
città d'arte	5
campagna	3
Totale complessivo	50

Modulo 1: elementi di statistica descrittiva

- **Terminologia essenziale**
- **Tipologie di variabili**
- **Scale**

Terminologia essenziale

1. unità statistica
2. collettivo statistico (*popolazione/campione*)
3. carattere
4. modalità

Terminologia essenziale (1)

Unità statistica

- L'oggetto dell'osservazione di ogni fenomeno individuale che costituisce il fenomeno collettivo è detto **unità statistica**.

L'unità statistica è l'unità elementare su cui viene osservato il fenomeno oggetto di studio.

Per esempio, è un'unità statistica:

ogni laureato residente in Italia di età compresa fra i 24 e i 35 anni;

ogni insegnante;

ogni piccola e media impresa abruzzese;

ogni processo civile etc..

Terminologia essenziale (2)

Collettivo statistico

L'insieme delle unità statistiche uguali rispetto ad alcune loro caratteristiche costituiscono il **collettivo statistico**

- Nel caso di indagine statistica, il collettivo può comprendere tutte le unità omogenee rispetto alle caratteristiche prescelte (**POPOLAZIONE**) o solo una parte di esse (**CAMPIONE**).

Per esempio: un campione di laureati residenti in Italia di età compresa fra i 24 e i 35 anni; l'insieme degli insegnanti della scuola primaria; un campione di piccole e medie imprese abruzzesi; un campione di alunni stranieri

Terminologia essenziale (2)

Popolazione e campione

Popolazione. Un insieme di elementi (persone fisiche, animali, squadre di calcio, visite ad un sito web), che è oggetto di studio statistico. Gli elementi della popolazione sono chiamati genericamente unità statistiche.

Campione. Un qualsiasi sotto-insieme della popolazione di riferimento. Il numero di unità statistiche presenti nel campione si chiama numerosità campionaria.

Una popolazione è detta reale quando è effettivamente esistente e visibile, mentre viene detta virtuale se è astratta o riferita al futuro.

Esempio 1. Gli insegnanti dei licei abruzzesi sono una popolazione reale.

Esempio 2. I pazienti affetti da una certa patologia (passati e futuri) sono una popolazione virtuale

Terminologia essenziale (3)

Carattere

Si chiama **carattere** ogni aspetto elementare oggetto di rilevazione sulle unità statistiche del collettivo.

- Ogni unità statistica è portatrice, generalmente, di molteplici caratteri.

- *Si consideri il collettivo degli studenti laureati stranieri.*

Per ognuno di essi si può rilevare l'età, il genere, il tipo di laurea conseguita, il luogo di residenza e la condizione occupazionale.

Terminologia essenziale (4)

Modalità

Si chiamano **modalità** di un carattere i diversi modi con cui questo si presenta nelle unità statistiche del collettivo

Ad esempio,

- ▣ *il genere può presentare modalità maschio/femmina;*
- ▣ *la condizione occupazionale può essere studente, lavoratore, in cerca di lavoro, disoccupato etc.*

Classificazione dei caratteri

I caratteri presenti in una unità statistica sono generalmente di natura diversa.

▣ **Caratteri qualitativi:** le modalità sono espressioni verbali

▣ *(es. genere, titolo di studio, anno di corso).*

▣ **Caratteri quantitativi:** le modalità possono essere espressi da numeri.
I caratteri quantitativi sono misurabili ed è possibile definire un'unità di misura

▣ *(es. età, reddito, altezza)*

Caratteri qualitativi

I caratteri qualitativi, a loro volta, possono distinguersi in sconnessi oppure ordinabili a seconda che sia possibile o meno graduarne la diversità.

■ **Caratteri qualitativi sconnessi:** hanno per modalità denominazioni verbali per le quali non esiste (e non è possibile stabilire) un ordinamento

■ *(genere, religione, regione di residenza, luogo di nascita, tipo di laurea).*

■ **Caratteri qualitativi ordinabili:** hanno per modalità denominazioni verbali per le quali esiste un ordinamento naturale

■ *(titolo di studio, anno di iscrizione all'università, grado di soddisfazione, livello di rischio associato ad un titolo azionario).*

Caratteri quantitativi

I caratteri quantitativi, a loro volta, possono distinguersi in discreti oppure continui.

■ **Caratteri quantitativi discreti**: le modalità sono i numeri interi 0, 1, 2, ...

- *(numero alunni di una classe, numero di addetti di un'azienda, numero di componenti di una famiglia, numero di chiamate in arrivo ad un call center).*

■ **Caratteri quantitativi continui**: le modalità sono tutti i numeri reali compresi in un determinato intervallo

- *(punteggio ad un test di ingresso, altezza, peso, tempo di attesa, durata di una conversazione telefonica, reddito).*

Operazioni che è possibile fare sulle modalità dei diversi tipi di carattere

		Operazioni		
		Uguaglianza/ Disuguaglianza	Ordinamento	Operazioni aritmetiche
Caratteri		$=, \neq$	$>, <$	$+, -, *, /$
Qualitativi	sconnessi	si	no	no
	ordinabili	si	si	no
Quantitativi		si	si	si

Quali sono

1.le unità statistiche?

2.le variabili?

3.le modalità?

**A che tipologia
appartengono i
seguenti caratteri:**

1. TIPOLOGIA DI INSEGNANTE

2. ETÀ

3. ANNI DI SERVIZIO

4. ORDINE SCUOLA IN CUI INSEGNA

5. GENERE

6. LA DISCIPLINA CHE HO STUDIATO È
STATA DECISIVA NELLA MIA SCELTA
DELLA PROFESSIONE DI
INSEGNANTE.

7. POSSIEDO UNA SOLIDA
CONOSCENZA DELLA MIA
DISCIPLINA, PER IL MIO LAVORO DI
INSEGNANTE.

ID	TIPOLOGIA DI INSEGNANTE	ETA'	ANNI DI SERVIZIO	ORDINE SCUOLA IN CUI INSEGNA	GENERE	La disciplina che ho studiato è stata decisiva nella mia scelta della professione di insegnante.	Possiedo una solida conoscenza della mia disciplina, per il mio lavoro di insegnante.
1	tutor di un neoassunto	41	11	Primaria	femminile	Molto d'accordo	D'accordo
2	tutor di un neoassunto	51	23	Secondaria di secondo grado	maschile	Molto d'accordo	Molto d'accordo
3	tutor di un neoassunto	34	5	Secondaria di primo grado	maschile	Molto d'accordo	D'accordo
4	tutor di un neoassunto	48	8	Secondaria di secondo grado	maschile	D'accordo	D'accordo
5	tutor di un neoassunto	45	10	Secondaria di secondo grado	maschile	Molto d'accordo	D'accordo
6	tutor di un neoassunto	31	6	Secondaria di secondo grado	maschile	Molto d'accordo	Molto d'accordo
7	tutor di un neoassunto	56	30	Secondaria di primo grado	femminile	Molto d'accordo	D'accordo
8	tutor di un neoassunto	46	19	Secondaria di secondo grado	femminile	Molto d'accordo	Molto d'accordo
9	tutor di un neoassunto	47	16	Secondaria di primo grado	femminile	Molto d'accordo	Molto d'accordo
10	tutor di un neoassunto	40	11	Primaria	femminile	D'accordo	Molto d'accordo
11	tutor di un neoassunto	61	40	Secondaria di primo grado	femminile	Abbastanza d'accordo	D'accordo
12	tutor di un neoassunto	55	26	Secondaria di secondo grado	maschile	Abbastanza d'accordo	D'accordo
13	tutor di un neoassunto	64	39	Secondaria di secondo grado	femminile	Molto d'accordo	Molto d'accordo
14	tutor di un neoassunto	53	15	Primaria	femminile	D'accordo	D'accordo

Classificazione alternativa (1): scale non metriche

➤ **SCALA NOMINALE**: il confronto fra due modalità del carattere è basato solo sulla relazione di equivalenza per cui si può dire solo se le modalità sono uguali o differenti (operazione di uguaglianza/disuguaglianza - *carattere qualitativo sconnesso*).

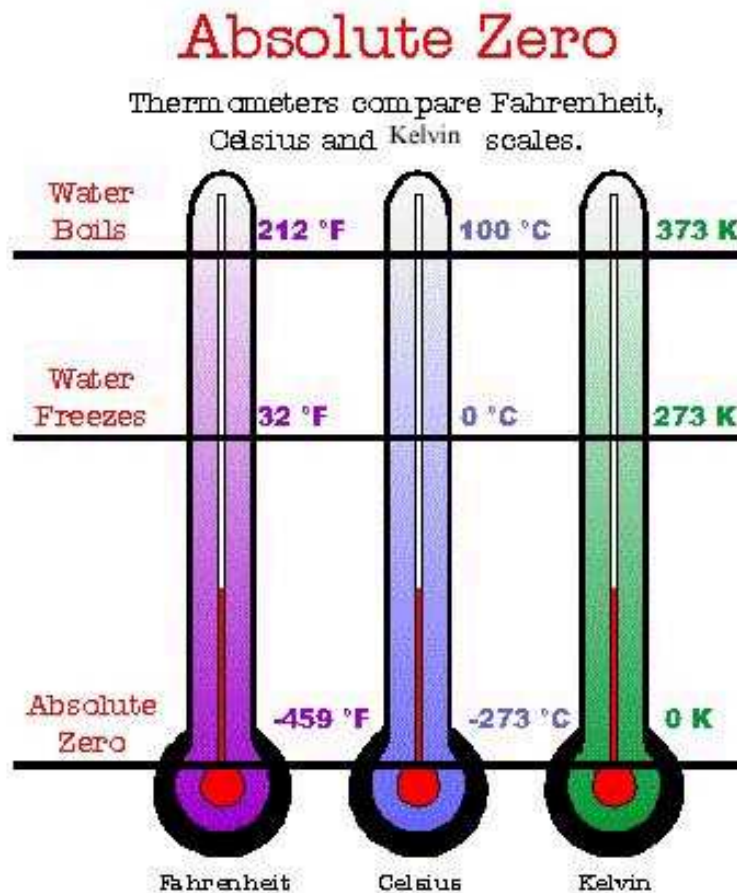
➤ **SCALA ORDINALE**: il confronto fra due modalità del carattere è basato sulla relazione d'ordine per cui oltre a dire se le modalità sono uguali o differenti si può anche dire qual è maggiore/minore (operazioni di uguaglianza/disuguaglianza e di ordinamento- *carattere qualitativo ordinabile*).

Classificazione alternativa (2): scale metriche

➤ **SCALA AD INTERVALLI**: il confronto fra due modalità dello stesso carattere può essere basato sulla loro differenza (carattere quantitativo per cui non esiste un'origine naturale. Lo zero non identifica l'assenza assoluta della caratteristica misurata - «zero relativo». *es. temperatura in gradi Celsius o Fahrenheit.*).

➤ **SCALA A RAPPORTI**: il confronto fra due modalità dello stesso carattere può essere basato sulla loro differenza e sul rapporto (carattere quantitativo per cui esiste un origine non arbitraria. Nella scala di rapporti il punto zero corrisponde all'assenza della proprietà oggetto di misurazione - «zero assoluto». *es. reddito, tasso di disoccupazione.*)

Temperatura: Scala ad intervalli



Nel caso della scala centigrada, per esempio, si è stabilito arbitrariamente che lo zero della scala corrisponde alla temperatura che porta al congelamento dell'acqua.

L'affermazione: «oggi è il doppio più caldo di ieri, dato che la temperatura di ieri era di 20° centigradi e la temperatura di oggi è di 40° centigradi» è priva di senso dato che se avessimo misurato queste medesime temperature usando la scala Fahrenheit, per esempio, i due valori sarebbero stati 68° F e 104° F.

LIVELLO DI MISURAZIONE	VARIABILE	MODALITA'	PROPRIETA' FORMALI
Scala nominale	Tipo di pensione posseduta	-nessuna -pensione sociale -pensione di anzianità/vecchiaia -pensione di reversibilità	Non è possibile stabilire intensità né distanze tra intervalli, ma solo che esistono delle differenze tra coloro che si collocano nell'una o nell'altra categoria
Scala ordinale	Giudizio sulla qualità del servizio domiciliare	-qualità pessima -qualità cattiva -qualità sufficiente -qualità discreta -qualità ottima	E' possibile stabilire l'intensità del giudizio ma senza disporre di una precisa distanza tra un livello di giudizio e l'altro
Scala a intervalli	Atteggiamento nei riguardi della religione	-livello 0 -livello 5 -livello 10 -livello 15 -livello 20 -livello 25	E' possibile stabilire l'uguaglianza degli intervalli che separano i diversi livelli, ma il punto zero è arbitrario.
Scala a rapporti	Reddito mensile	-nullo - fino a 250 euro - da 250 a 500 euro - da 500 a 1000 euro - da 1000 a 2000 euro - oltre 2000 euro	E' possibile stabilire il reddito di ogni rispondente e l'ampiezza dell'intervallo che lo separa da un altro. Vi è un punto zero che consiste nell'assenza di reddito.

Sentirete parlare anche di...

Si distinguono le variabili qualitative in base al numero di modalità che possono assumere:

Variabili dicotomiche (con due modalità di risposta come: sì o no, presente o assente, ecc.).

Variabili politomiche (che possono assumere più di due modalità di risposta come corso di laurea, titolo di studio, ecc.).

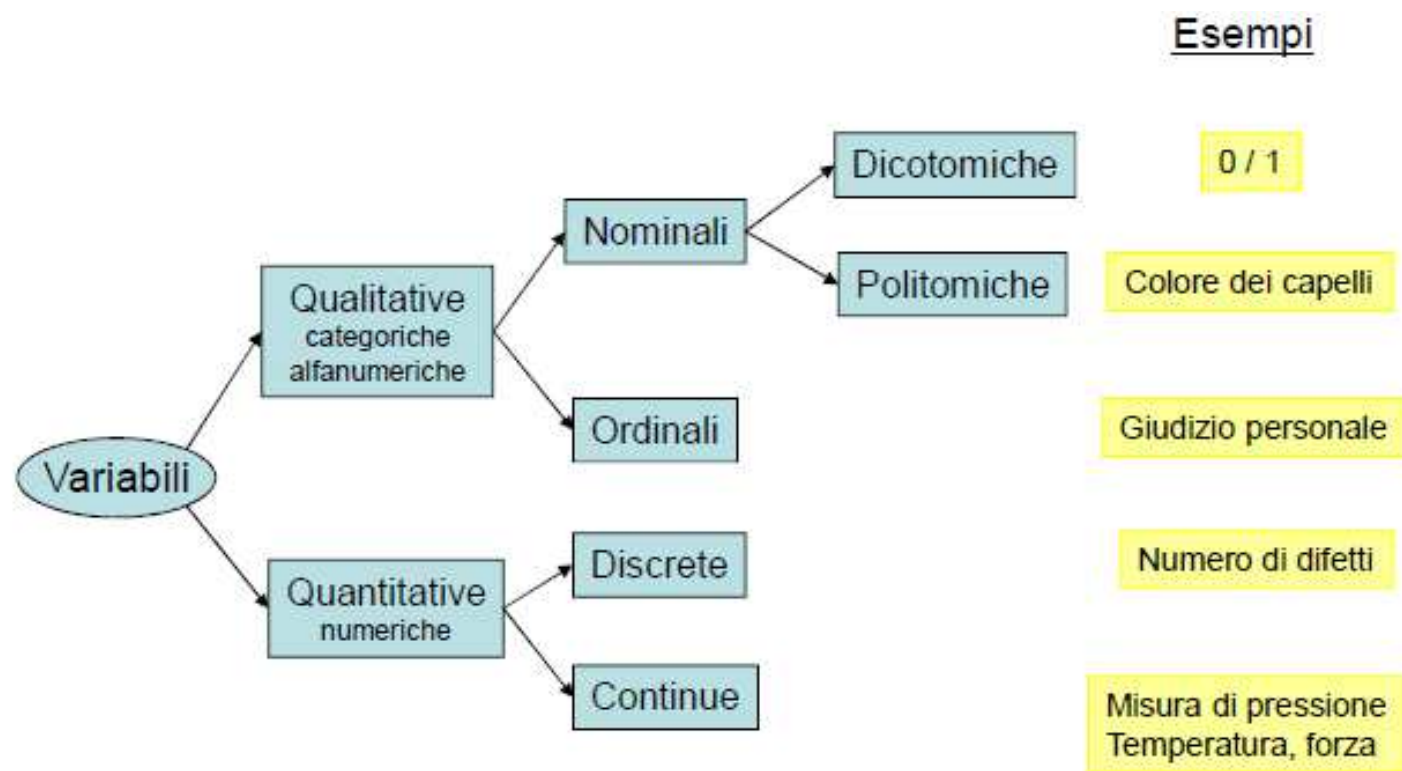
Possiamo distinguere anche le variabili in base al ruolo assunto nella spiegazione dei fenomeni in esame.

Possiamo ad esempio assumere una dipendenza tra due variabili, individuandone una come causa e l'altra come effetto.

Variabili indipendenti: la variabile individuata come causa (ad es. titolo di studio).

Variabili dipendenti. la variabile individuata come effetto (ad es. reddito mensile netto).

In sintesi...



Statistica e Valutazione della Formazione — Slides 1

Usa queste domande per verificare la tua comprensione degli argomenti trattati.
Prima di guardare le risposte, prova a rispondere autonomamente.

Definizione e natura della statistica

- 1 Qual è la differenza tra la statistica come «scienza dello Stato» e la statistica moderna?
- 2 In che senso la statistica è una «scienza ancillare»? Fai due esempi in ambito educativo.
- 3 Cosa si intende per Data Science e in che rapporto sta con la statistica?
- 4 Secondo Domenico Piccolo, la statistica è «la scienza delle decisioni in condizioni di incertezza». Cosa significa questo per un pedagogo?
- 5 Qual è la differenza tra statistica descrittiva e statistica inferenziale?

Terminologia essenziale

- | | |
|----|--|
| 6 | Definisci «unità statistica» e fai un esempio riferito al contesto scolastico. |
| 7 | Qual è la differenza tra popolazione e campione? Cosa si intende per «popolazione virtuale»? |
| 8 | Cos'è un carattere statistico? E una modalità? Fai un esempio con il carattere «titolo di studio». |
| 9 | Classifica questi caratteri come qualitativi sconnessi, qualitativi ordinabili o quantitativi:
genere • anno di corso • numero di assenze • grado di soddisfazione • regione di residenza |
| 10 | Qual è la differenza tra carattere quantitativo discreto e continuo? Dai un esempio per ciascuno. |
| 11 | Cosa si intende per variabile dipendente e variabile indipendente? Fai un esempio in campo educativo. |



Scale di misura

12 Descrivi le quattro scale di misura (nominale, ordinale, a intervalli, a rapporti) e le operazioni permesse da ciascuna.

13 Perché i voti scolastici (da 1 a 10) non possono essere considerati una vera scala a intervalli? Che tipo di scala sono?

14 La temperatura in gradi Celsius è una scala a intervalli o a rapporti? Perché l'affermazione «oggi è il doppio più caldo di ieri» è priva di senso?

15 Classifica queste variabili per scala di misura:
stato civile • grado di soddisfazione • reddito mensile • temperatura corporea • anno di corso

16 Quale scala di misura consente di calcolare la media aritmetica? E quale invece no? Perché?

Mappa concettuale del Modulo 1

STATISTICA

Scienza dei dati e dell'incertezza
Descrittiva ↔ Inferenziale

TERMINOLOGIA

Unità · Collettivo · Carattere · Modalità
Popolazione ↔ Campione

TIPI DI CARATTERE

Qualitativo (sconnesso / ordinabile)
Quantitativo (discreto / continuo)

SCALE DI MISURA

Nominale → Ordinale
A intervalli → A rapporti

VARIABILI

Dipendente ↔ Indipendente
Dicotomica ↔ Politomica

AMBITO PEDAGOGICO

Fenomeni collettivi scolastici
Voti, DSA, dispersione, INVALSI