

GLOSSARIO

Un glossario di nozioni e concetti di statistica per insegnanti e dirigenti può essere un'ottima risorsa per chiunque voglia comprendere a fondo alcuni concetti della statistica, capire meglio il valore di dati e analisi, insegnare e trasferire tali concetti durante il proprio insegnamento.



1

POPOLAZIONE

L'insieme completo di tutti gli elementi che si stanno studiando. Ad esempio, se stiamo studiando l'altezza degli studenti di una scuola, la popolazione sarebbero tutti gli studenti della scuola.

2

CAMPIONE

Un sottoinsieme rappresentativo della popolazione. Spesso, a causa di limiti di tempo o risorse, non è pratico studiare un'intera popolazione.

Quindi, selezioniamo un campione che rifletta accuratamente la popolazione.

3

CAMPIONAMENTO CASUALE

Un metodo di selezione del campione in cui ogni elemento della popolazione ha la stessa probabilità di essere incluso nel campione. Il campionamento casuale è importante per garantire la rappresentatività del campione rispetto alla popolazione.

4

OPERATIVIZZAZIONE

Processo per tradurre concetti astratti in concetti empirici misurabili e osservabili. Consiste nella identificazione di indicatori o misure che rappresentino in maniera efficace quel concetto. Per esempio, per misurare il "successo accademico", possiamo usare indicatori diversi: voto dello studente, risultato nel test standardizzato, tasso di frequenza scolastica, etc.

5

VARIABILE

Una caratteristica o un attributo che può assumere diversi valori. Ad esempio, l'altezza degli studenti è una variabile.

6

VARIABILE DIPENDENTE

La variabile che viene misurata o osservata in uno studio. È influenzata dalle altre variabili nello studio. Ad esempio, se stiamo studiando l'effetto dello studio sulla performance degli studenti, la performance degli studenti sarebbe la variabile dipendente.

7

VARIABILE INDIPENDENTE

La variabile che il ricercatore manipola o controlla. È la causa presunta della variazione nella variabile dipendente. Utilizzando l'esempio precedente, lo studio sarebbe la variabile indipendente.

8

DISTRIBUZIONE

La maniera in cui i valori di una variabile sono distribuiti. Ad esempio, la distribuzione delle altezze degli studenti potrebbe essere normale o uniforme.

9

DISTRIBUZIONE NORMALE

Una distribuzione simmetrica di dati, dove la maggior parte dei valori si concentra intorno alla media e la forma della distribuzione è definita da parametri come la media e la deviazione standard.

10

MODA

Il valore più frequente in una distribuzione. Ad esempio, in {2, 3, 3, 5, 7}, la moda è 3, perché è il numero che compare più volte.

11

MEDIANA

Il valore centrale di una distribuzione ordinata.

Metà dei valori sono al di sopra e metà al di sotto di questo valore. Ad esempio, in {2, 3, 3, 5, 7}, la mediana è 3, perché è il numero al centro della serie ordinata.

12

MEDIA

Il valore medio di una distribuzione, calcolato sommando tutti i valori e dividendo per il numero totale di valori. È spesso indicato come il valore "centrale" di una distribuzione. Per esempio, in {2, 3, 3, 5, 7}, la media è $(2 + 3 + 3 + 5 + 7) / 5 = 4$.

13

DEVIAZIONE STANDARD

Una misura di dispersione che indica quanto i valori di una distribuzione si discostano dalla media.

Una deviazione standard maggiore indica una maggiore dispersione dei dati.

14

ERRORE STANDARD

Una misura della precisione con cui una stima di un parametro statistico si avvicina al suo valore vero. L'errore standard è spesso utilizzato per calcolare gli intervalli di confidenza intorno alle stime dei parametri.

15

INTERVALLO DI CONFIDENZA

Quando stimiamo un parametro di una popolazione, l'intervallo di confidenza indica un'ampiezza delimitata da due limiti (inferiore e superiore) entro cui si colloca la nostra stima del parametro, con un certo grado di probabilità. L'intervallo di confidenza diminuisce se diminuisce la variabilità nella popolazione (errore standard) e aumenta se aumenta la numerosità del campione.

16

REGOLA EMPIRICA (O REGOLA DEL 68-95-99.7)

Una regola approssimativa che descrive la distribuzione dei valori in una distribuzione normale. Secondo questa regola, circa il 68% dei dati si trova entro una deviazione standard dalla media, circa il 95% entro due deviazioni standard e circa il 99.7% entro tre deviazioni standard.

17

P-VALUE

Un valore numerico che descrive la probabilità di osservare un risultato che non è dettato dal caso, ma che assume valore statistico significativo. Un valore basso suggerisce che le nostre ipotesi sulla relazione tra due variabili non sono sbagliate.

18

CORRELAZIONE

Una misura della relazione tra due variabili. Può essere positiva (entrambe le variabili aumentano insieme), negativa (una variabile aumenta mentre l'altra diminuisce) o neutra (nessuna relazione).

19

PREDIZIONE

Un'istanza di utilizzo di modelli statistici o algoritmi per stimare o prevedere il valore di una variabile in base ai valori delle variabili predictorie. Le predizioni sono spesso utilizzate per fare ipotesi su eventi futuri o per riassumere modelli complessi in modo più semplice.

20

PROFILAZIONE

Il processo di utilizzo dei dati per identificare e descrivere le caratteristiche di un gruppo di individui o oggetti. La profilazione può essere utilizzata per identificare tendenze, cluster o segmenti all'interno di una popolazione.

21

CLUSTER ANALYSIS

Una tecnica statistica utilizzata per identificare gruppi omogenei all'interno di un insieme di dati, in base alle somiglianze tra gli elementi.

22

REGRESSIONE

Una tecnica statistica utilizzata per studiare la relazione tra una variabile dipendente e una o più variabili indipendenti. La regressione può essere utilizzata per fare predizioni o per comprendere meglio la natura delle relazioni tra le variabili.

23

OUTLIER

Un valore che si discosta significativamente dagli altri valori in un insieme di dati. Gli outlier possono influenzare negativamente i risultati dell'analisi statistica e devono essere gestiti o trattati di conseguenza.

24

PROBABILITÀ

La probabilità è la proporzione di volte in cui una possibilità di un evento si verifica se osserviamo un processo randomico per un numero infinito di volte (legge dei grandi numeri). Per esempio, se lanciamo una moneta un infinito numero di volte, sappiamo che la probabilità che esca testa è 50% e la probabilità che esca croce è 50%.

25

PROBABILTÀ SCOMPOSTA

La probabilità scomposta permette di calcolare la probabilità di un evento dividendo lo spazio in scenari separati e sommando le probabilità di ciascuno scenario ponderate dalla loro probabilità.

E' dunque come suddividere un problema complesso in parti più semplici, calcolare le probabilità per ciascuna parte e poi sommarle.

Ad esempio, se si vuole sapere la probabilità che piova oggi, è possibile considerare la probabilità che piova al mattino e al pomeriggio separatamente e poi sommarle.



LA DIFFERENZA PRINCIPALE TRA DATO E INFORMAZIONE RISIEDE NEL LORO SIGNIFICATO E NEL LORO CONTESTO DI UTILIZZO:

1

DATO

Un dato è una rappresentazione grezza, non elaborata o non elaborata di fatti, numeri, misurazioni o descrizioni. È una singola unità di informazione che, per sé stessa, potrebbe non avere un significato particolare. Ad esempio, "10", "rosso", "2,5 kg" sono tutti esempi di dati. I dati possono essere raccolti, registrati o ottenuti da varie fonti come misurazioni, osservazioni, sondaggi, sensori, ecc.

2

INFORMAZIONE

L'informazione, d'altra parte, è il risultato della manipolazione, dell'elaborazione o dell'interpretazione dei dati, che dà loro significato o contesto. L'informazione è una raccolta organizzata e significativa di dati che trasmette un messaggio o aggiunge conoscenza. Ad esempio, se diciamo che "la temperatura è di 10 gradi Celsius", questo dato acquisisce significato solo quando viene interpretato nel contesto di una previsione meteorologica o di un contesto specifico.

In breve, i dati sono i mattoni grezzi della conoscenza, mentre l'informazione è il prodotto elaborato e significativo che emerge dalla manipolazione o dall'analisi dei dati. Quindi, mentre i dati sono il materiale di base, l'informazione è ciò che trasmette valore e comprensione.



I DATI POSSONO ESSERE DI DIVERSI TIPI, A SECONDA DELLA NATURA DELLE INFORMAZIONI CHE RAPPRESENTANO

1

DATABASE

E' un insieme organizzato di dati, in genere archiviato elettronicamente in un sistema informatico e strutturato in una matrice, dove ogni singola riga corrisponde ad un caso o a un'osservazione (esempio, studente A, studente B, studente C), mentre ogni colonna corrisponde a una variabile (esempio, età, genere, voto in matematica, voto in italiano...)

2

VARIABILE

Una variabile è un elemento, una caratteristica o una proprietà che può assumere valori diversi per ogni caso osservato. Le variabili sono utilizzate nelle indagini statistiche per raccogliere dati e possono rappresentare caratteristiche diverse.

CASI/
OSSERVAZIONI

ID STUDENTE	ANNO DI NASCITA	GENERE	VOTO IN MATEMATICA	VOTO IN ITALIANO
Studente A	2003	M	5	8
Studente B	2007	F	7	7
Studente C	2005	F	6	8
Studente D	2006	M	9	8
...	...	...	...	...
...	...	...	...	...
Studente Z	2005	F	7	6

VARIABILI

ALCUNI PRINCIPALI TIPI DI VARIABILI

1

VARIABILI NOMINALI O CATEGORIALI

Sono variabili che rappresentano categorie o gruppi distinti. I valori delle variabili nominali non hanno un ordine intrinseco che ha un significato

Esempio: colore degli occhi (blu, verde, marrone)

2

VARIABILI ORDINALI

Sono variabili che rappresentano categorie con un ordine intrinseco, e i cui valori possono essere ordinati secondo una logica. Tuttavia, la distanza tra le categorie non è necessariamente uguale.

Esempio: livello di istruzione (elementare, medio, superiore, universitario); grado di soddisfazione (molto insoddisfatto, insoddisfatto, neutro, soddisfatto, molto soddisfatto)

3

VARIABILI CARDINALI (O QUANTITATIVE)

Sono variabili che rappresentano quantità numeriche e sono suddivise in:

-Variabili discrete: possono assumere solo valori specifici, spesso numero interi e non valori intermedi

esempio: numero di figli, numero di libri posseduti

-Variabili continue: possono assumere un qualsiasi valori all'interno di un intervallo e possono essere frazionari

Esempio: altezza, peso

4

VARIABILI BINOMIALI O DICOTOMICHE

Sono variabili che possono assumere solo due valori possibili

Esempio: risposta si/no, vero/falso.

LE VARIABILI DELL'ISTRUZIONE

Nel mondo dell'istruzione, troviamo delle variabili ricorrenti che sono fondamentali per lo studio delle dinamiche scolastiche. E' importante conoscerle per poterle misurare nel modo corretto, usando gli indicatori giusti. Un giusto processo di operativizzazione permette di ottenere sempre delle misure valide (ovvero capaci di catturare il concetto desiderato in maniera accurata) e affidabili (ovvero capaci di misurare il concetto desiderato in maniera consistente ad ogni misurazione).

A questo scopo, gli studiosi e gli esperti di educazione creano test, questionari e scale validati, ovvero che:

- 1. Contengono le domande giuste per il gruppo target della ricerca (età, condizioni specifiche)**
- 2. Contengono domande o items che corrispondono a dimensioni, aree o temi legati al macro-tema che si sta analizzando**
- 3. Permettono di ottenere una misura sempre comparabile nel tempo e tra diversi contesti**
- 4. Permettono di ottenere una misura sempre valida e affidabile**

AFFIDABILE MA
NON VALIDO



VALIDO MA
NON AFFIDABILE



AFFIDABILE
E VALIDO



ESEMPI DI TEST E SCALE VALIDATE

1

TEST STANDARDIZZATI

I test standardizzati sono test validati da esperti di istruzione, e sono delle prove effettuate e valutate in modo consistente per tutti i partecipanti (sono uguali per tutti). I test standardizzati sono progettati in modo tale che le domande, le condizioni, e l'attribuzione del punteggio siano sempre coerenti e non siano determinati da condizioni esterne dalla prova stessa. Nella maggior parte dei casi, i test standardizzati sono intesi come strumento di valutazione del sistema di educazione e formazione nel suo complesso: per questo motivo sono spesso somministrati su scala nazionale.

Esempio: In Italia, i test INVALSI permettono, attraverso la misurazione standardizzata degli esiti di apprendimento degli allievi, un'analisi territoriale e temporale delle criticità al fine di migliorare il sistema scolastico. L'obiettivo dei test INVALSI NON è la misurazione delle abilità per il confronto tra singoli studenti, classi o scuole. A differenza del voto, che in quanto tale include la valutazione da parte dell'insegnante di diverse dimensioni (capacità, miglioramento, impegno...), il test standardizzato viene considerato più «neutro», in quanto il suo scopo è permettere, attraverso un'analisi di contesto, l'individuazione di situazioni critiche al fine di intervenire per un miglioramento complessivo del nostro sistema di istruzione. E' importante sottolineare che i test standardizzati e la valutazione degli allievi da parte degli insegnanti alla fine o durante l'anno scolastico sono due misure molto diverse e con obiettivi diversi, e sono quindi complementari: l'una non può e non vuole sostituire l'altra.



La tecnica delle scale consiste in un insieme di procedure messe a punto da esperti per misurare concetti complessi e non direttamente osservabili. Una scala è un insieme coerente di elementi che sono considerati indicatori di un concetto generale. Usare una scala validata, ovvero creata e testata da esperti, permette di ottenere una misura puntuale, sempre comparabile nel tempo e tra diversi contesti.

Esempio: AMS (Academic Motivation Scale) è una scala usata per misurare la motivazione degli studenti nelle sue diverse dimensioni: regolazione intrinseca, regolazione identificata, regolazione introiettata, regolazione esterna, amotivazione.

AMS (Academic Motivation Scale): la scala di risposta è 1 = per niente; 2 = poco; 3 = abbastanza; 4 = molto

Gli item della scala

1. Perché voglio guadagnare di più di chi non ha studiato
2. Perché mi piace imparare cose nuove
3. Per prepararmi per quello che mi piacerebbe fare in futuro
4. Perché mi piace veramente andare a scuola per imparare
5. Sinceramente vado a scuola solo perché mi costringono.
6. Perché così posso far vedere che sono capace di essere promosso/a
7. Perché è bello scoprire cose che non conoscevo prima
8. Perché voglio avere un lavoro pagato molto bene
9. Per poter fare in futuro le cose che mi piacciono
10. Vado a scuola solo perché devo farlo
11. Perché quando sono capace di prendere buoni voti mi sento importante
12. Per prepararmi meglio per la scuola che mi piacerebbe fare poi
13. A dire la verità non me ne importa nulla di andare a scuola
14. Per far vedere che sono intelligente
15. Perché voglio diventare una persona importante, qualcuno che conti
16. Perché a scuola posso continuare a studiare cose che mi piacciono
17. Per imparare cose che mi serviranno per quello che poi mi piacerebbe fare nella vita
18. Sinceramente non lo so. Sento che sto solo perdendo tempo a scuola
19. Per far vedere che quando voglio vado bene a scuola
20. Perché voglio fare più soldi quando comincio a lavorare

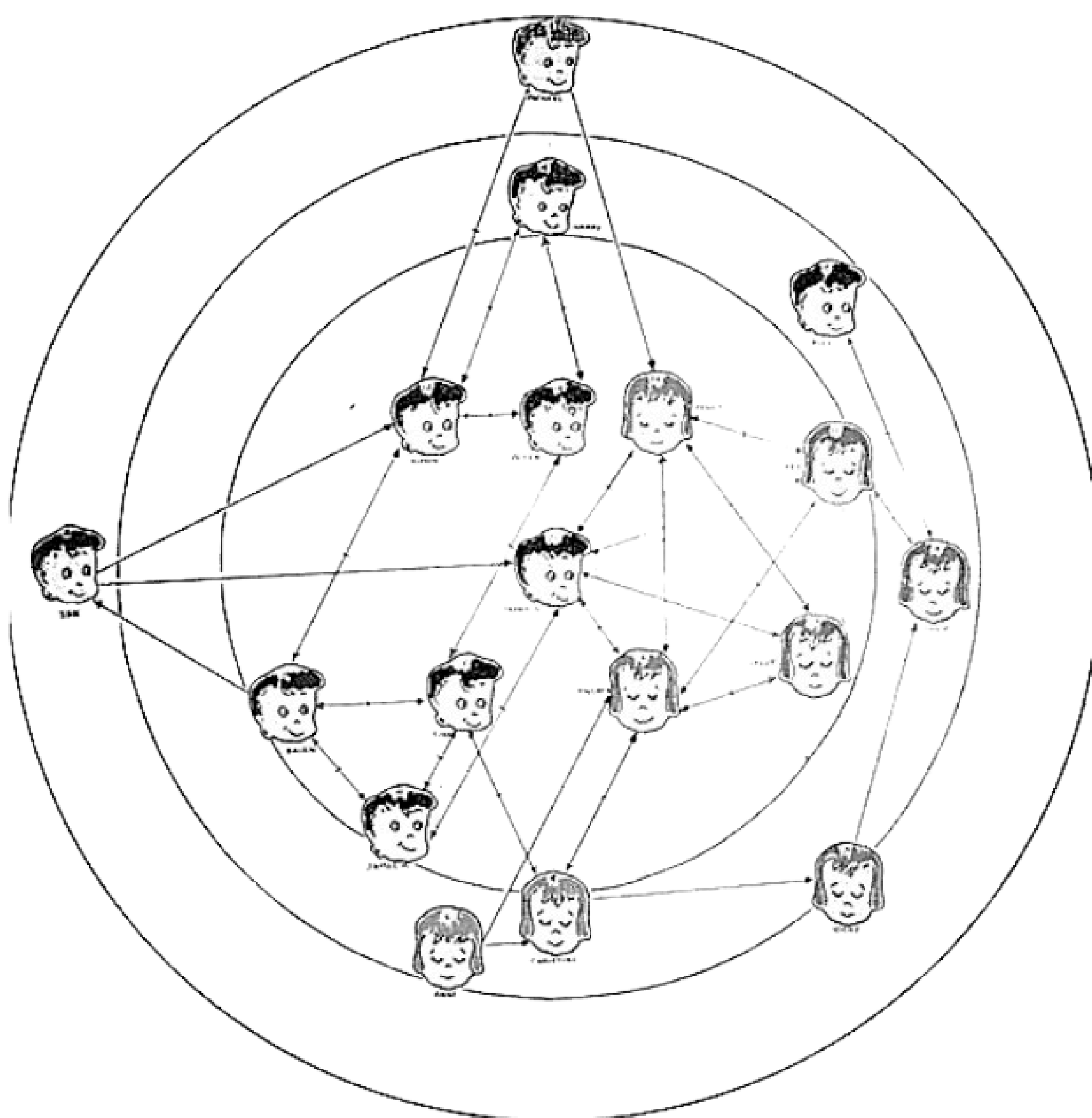
Il test sociometrico nasce per analizzare le relazioni interpersonali esistenti all'interno di un gruppo di individui; infatti il suo campo di applicazione ideale è rappresentato dalla classe scolastica. Il test sociometrico consiste in un questionario con poche domande, che ruotano intorno al tema della preferenza e rifiuto nei confronti degli altri appartenenti al gruppo. In questo modo si può conoscere il grado di prestigio di cui gode un elemento rispetto agli altri componenti del gruppo, e l'organizzazione interna del gruppo. E' un test utile sia come strumento di diagnosi individuale che per cogliere la struttura relazionale del gruppo.

Esempio di domande di test sociometrico:

1. Se dovessi scegliere uno dei membri del tuo gruppo per un lavoro particolarmente impegnativo chi sceglieresti? Scelta n. 1 (NOME) Scelta n. 2 (NOME)

2. Se dovessi scegliere uno dei membri del tuo gruppo come tuo compagno di gita, chi sceglieresti? Scelta n. 1 (NOME) Scelta n. 2 (NOME)

3. Secondo una tua profonda convinzione, chi ti ha scelto come compagno di lavoro? Scelta n. 1 (NOME) Scelta n. 2 (NOME)



- Scala di Likert
 - Scalogramma di Guttman
 - Scala di Rasch (modelli probabilistici)
 - Differenziale semantico
- ... E molti altri.

I vantaggi dell'utilizzazione di scale e test validati da esperti è che sono sempre affidabili e validi. Il campo dell'istruzione è particolarmente ricco di test e scale validate per misurare moltissimi concetti diversi, sia a livello di studente (motivazione, empatia, felicità...) sia di classe o scuola (inclusione) e sono reperibili facilmente in internet. Gli svantaggi sono che spesso richiedono del tempo di preparazione e compilazione da parte dei partecipanti, e delle capacità di statistica di base per analizzare i risultati. Inoltre, i questionari e le scale esistenti già validate non possono essere personalizzati a seconda delle necessità specifiche (per esempio, adattando gli items o modificando l'ordine), e non sempre misurano esattamente il concetto che vorremmo analizzare.



RAPPRESENTAZIONI GRAFICHE DI DATI E VARIABILI

ESEMPI CON UNA VARIABILE

GRAFICO A BARRE

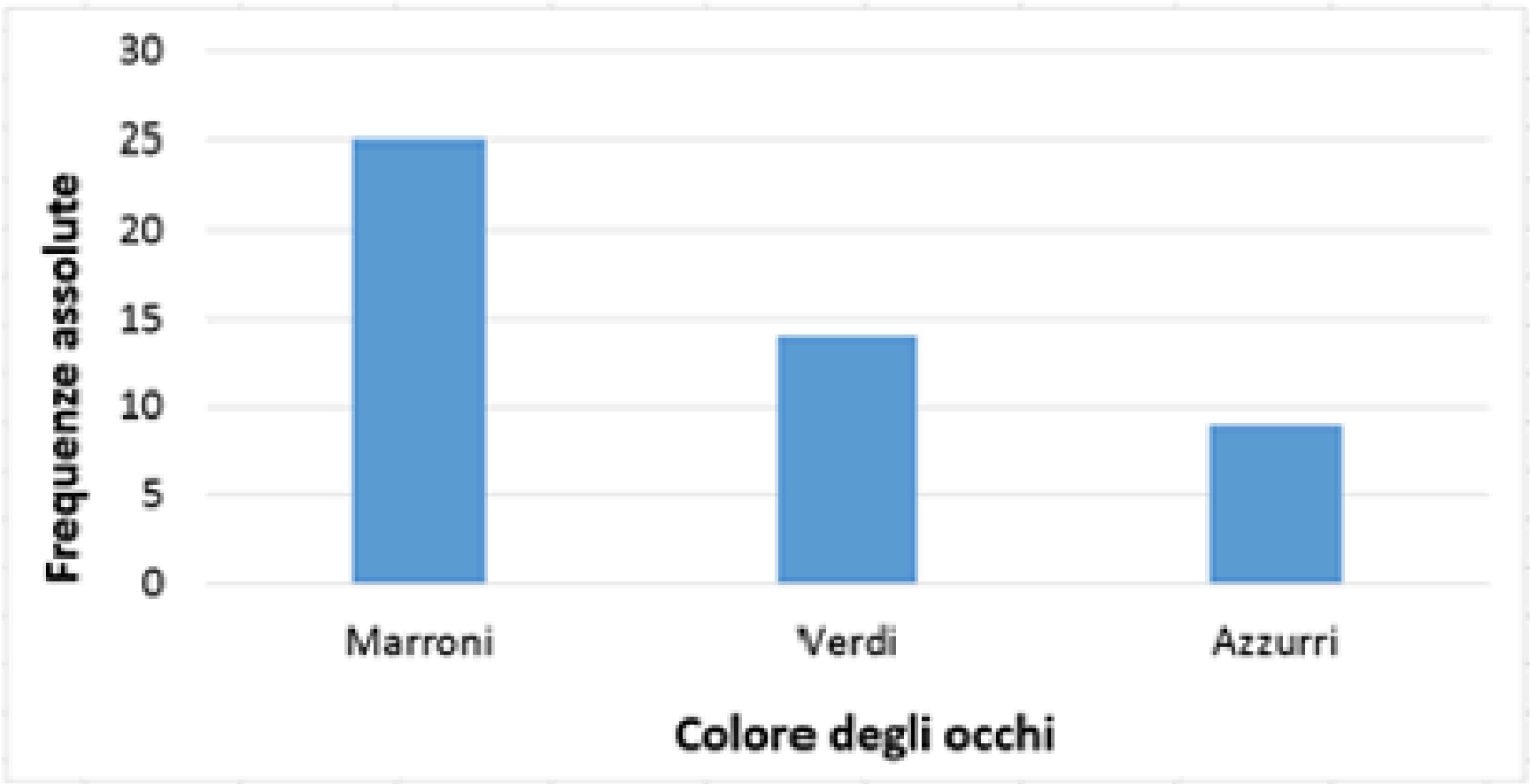


GRAFICO A TORTA

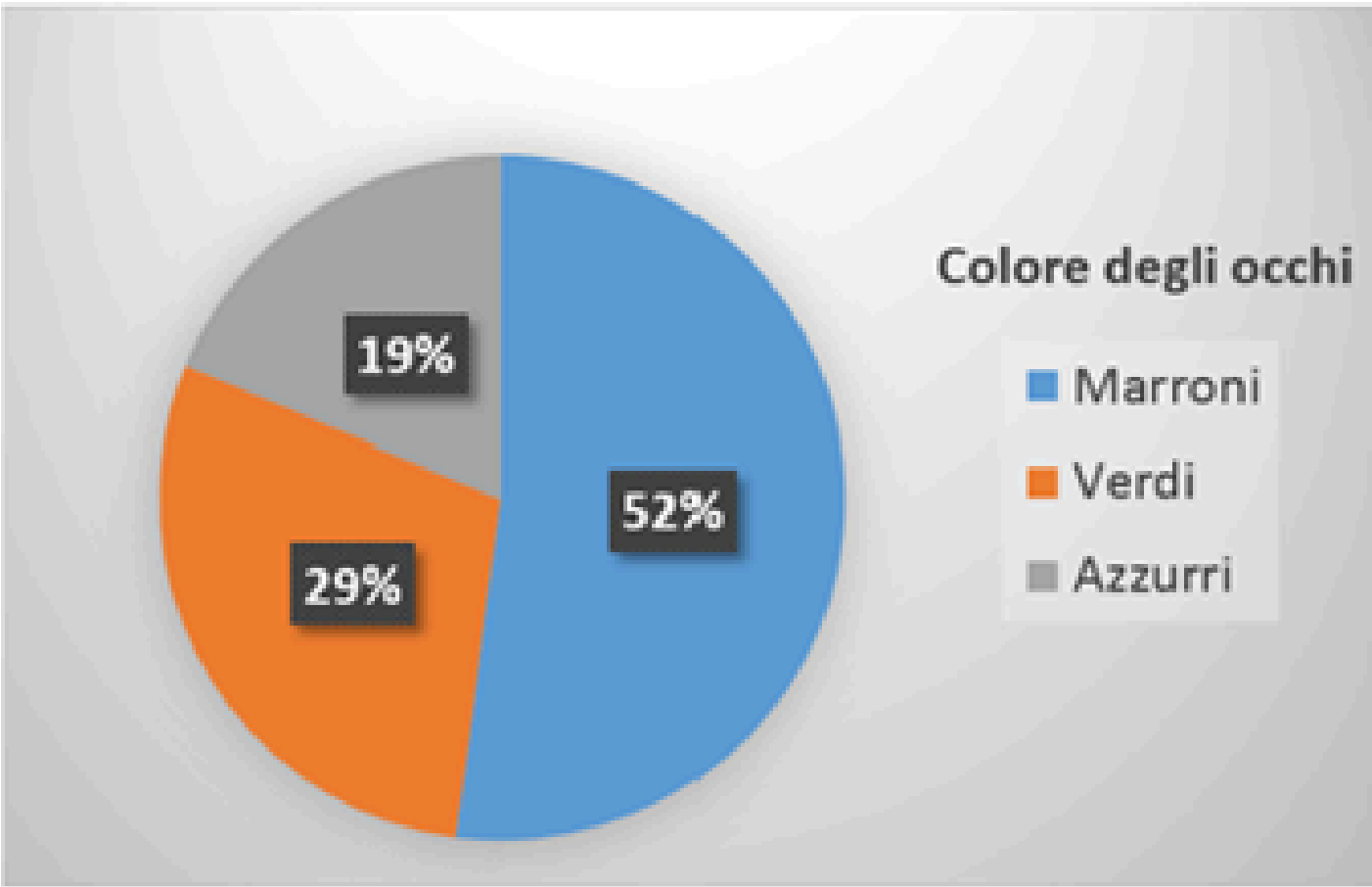
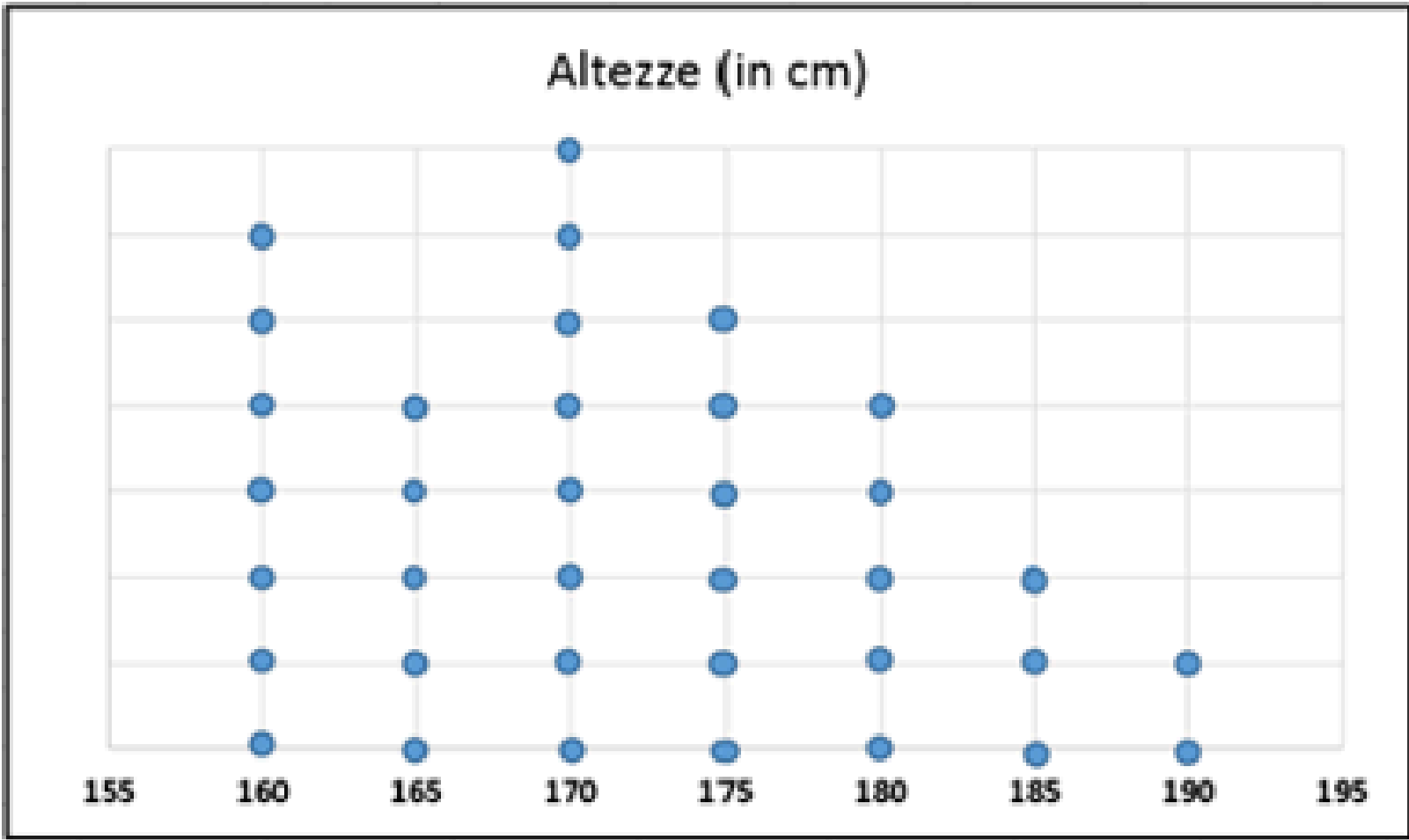
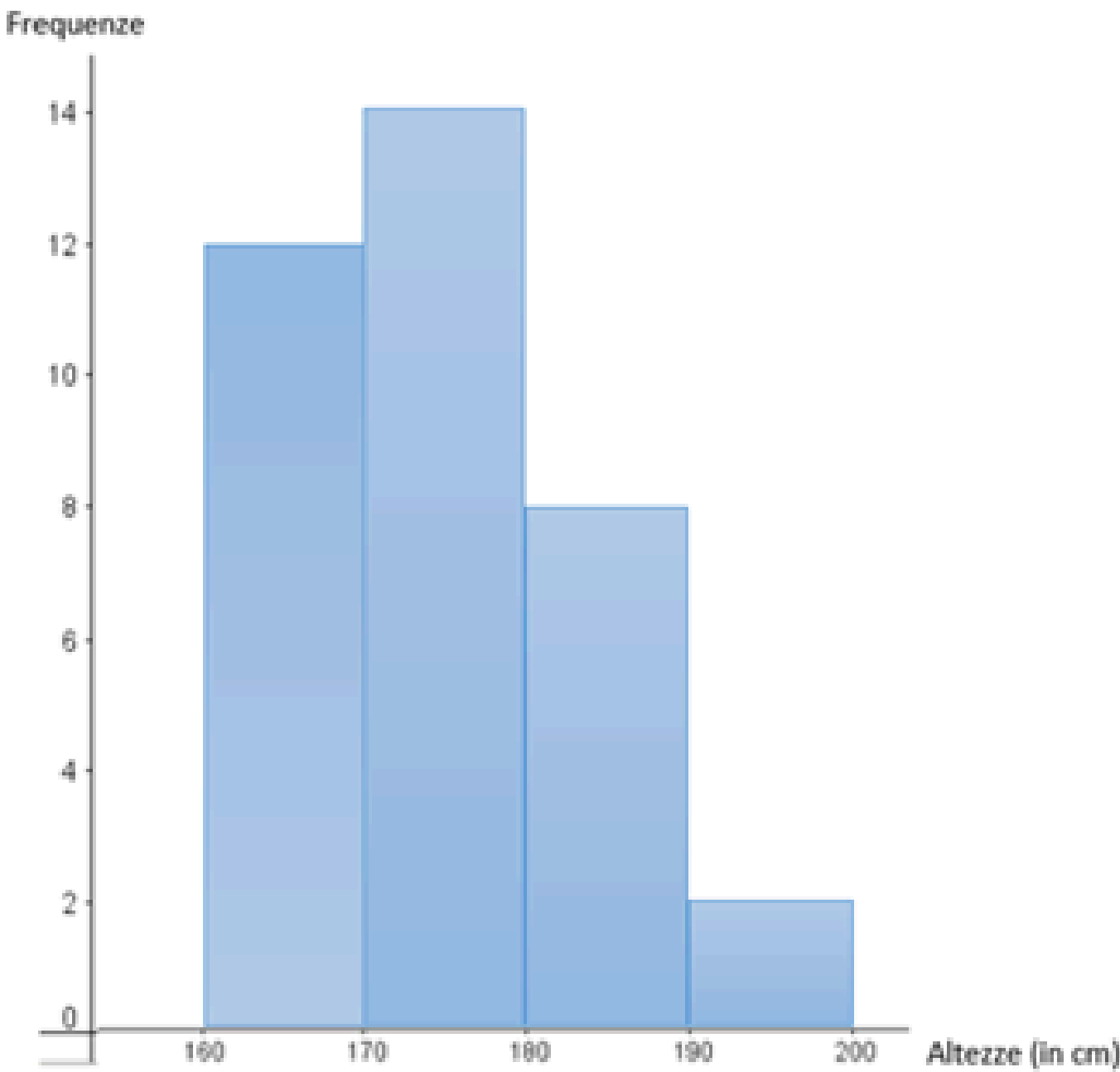


DIAGRAMMA A DISPERSIONE



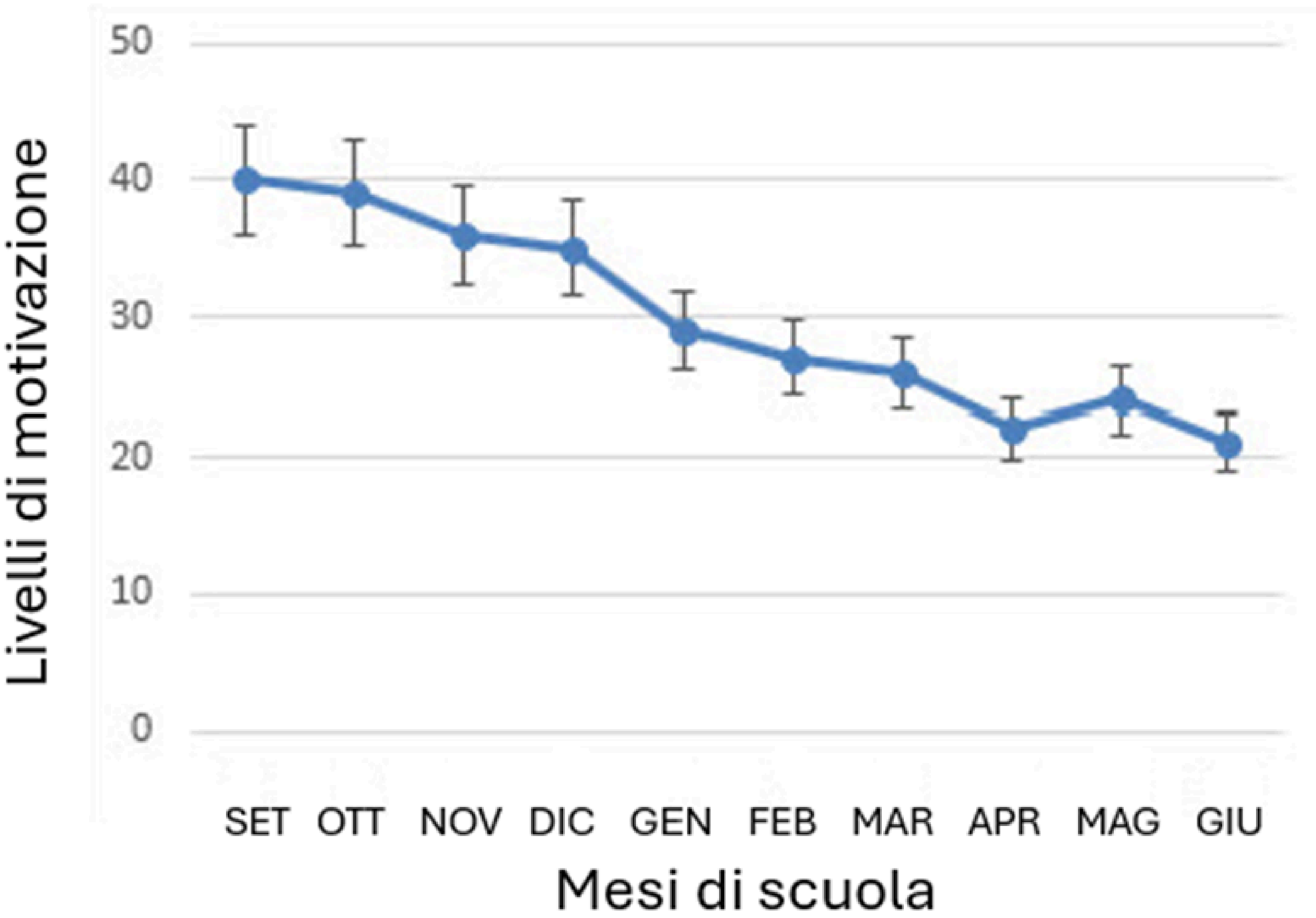
ISTOGRAMMA



ESEMPI CON DUE O PIÙ VARIABILI

1 INTERVALLI DI CONFIDENZA

Gli intervalli di confidenza indicano che non ci sono differenze statisticamente significative nei livelli di motivazione tra il primo e il secondo mese di scuola, o tra il primo e il terzo (gli intervalli si sovrappongono). Mentre il livello di motivazione cala significativamente confrontando il primo mese di scuola e il quinto, oppure il primo e 10 decimo.



2

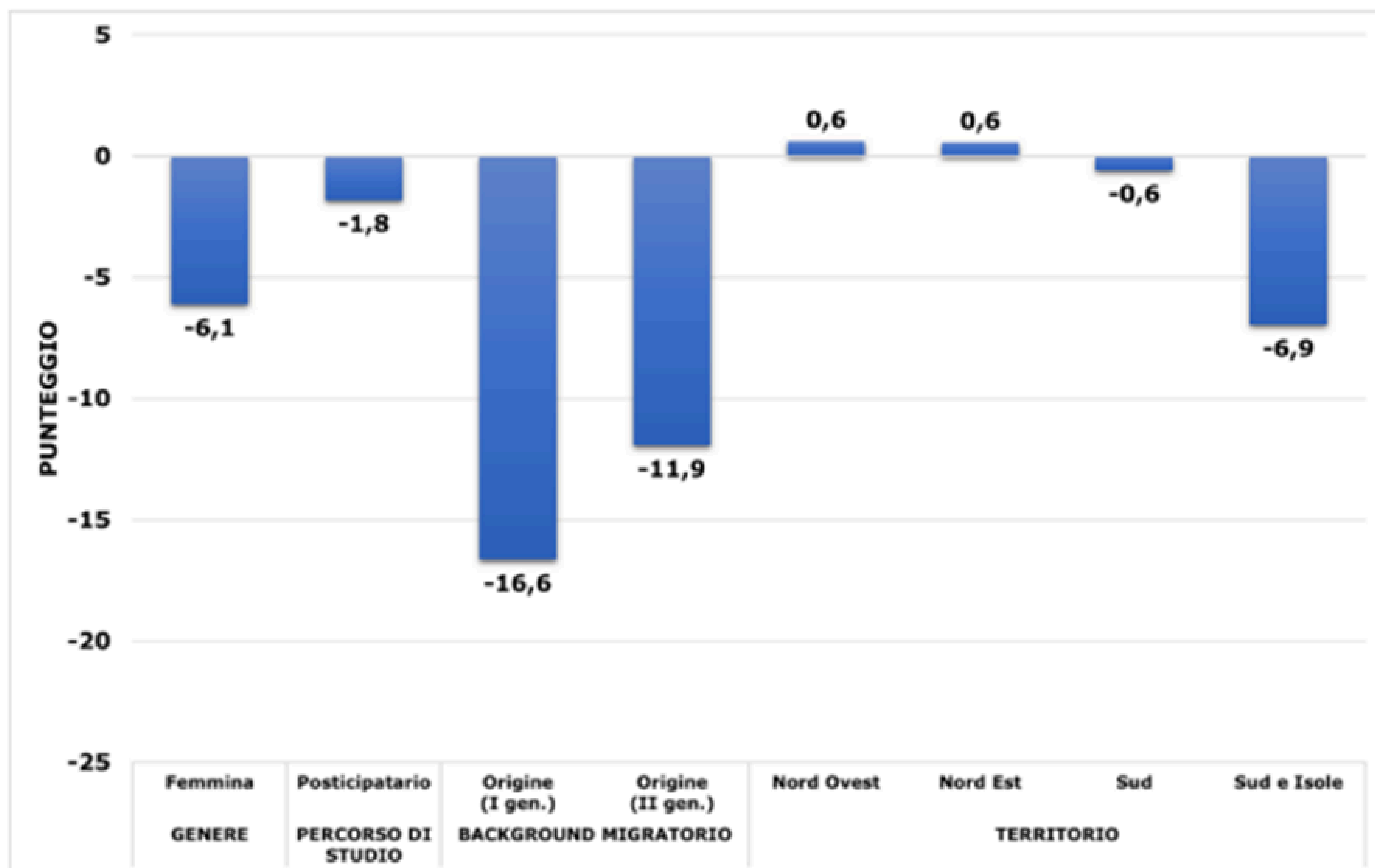
GRAFICO A BARRE

E' importante leggere i titoli dei grafici, che ci forniscono delle indicazioni su come leggere quello che viene rappresentato. In questo grafico, le colonne riportano differenze di punteggio percentuale: se sono in basso (sotto lo zero) le differenze sono negative per le categorie riportate nelle etichette.

Per esempio, le femmine (etichetta «genere») hanno un punteggio nella prova di Matematica in II primaria più basso di 6,1 punti percentuali in media rispetto ai maschi.

Gli studenti migranti (I generazione) hanno un punteggio in media inferior di 16,6 punti percentuali rispetto agli studenti italiani, mentre gli studenti migranti (II generazione) hanno un punteggio in media inferiore di 11,9 punti percentuali rispetto agli studenti italiani.

Figura 2.3.2.1 – Peso di alcuni fattori sull'esito della prova di Matematica in II primaria. Differenza di punteggio percentuale (fonte: INVALSI 2023)



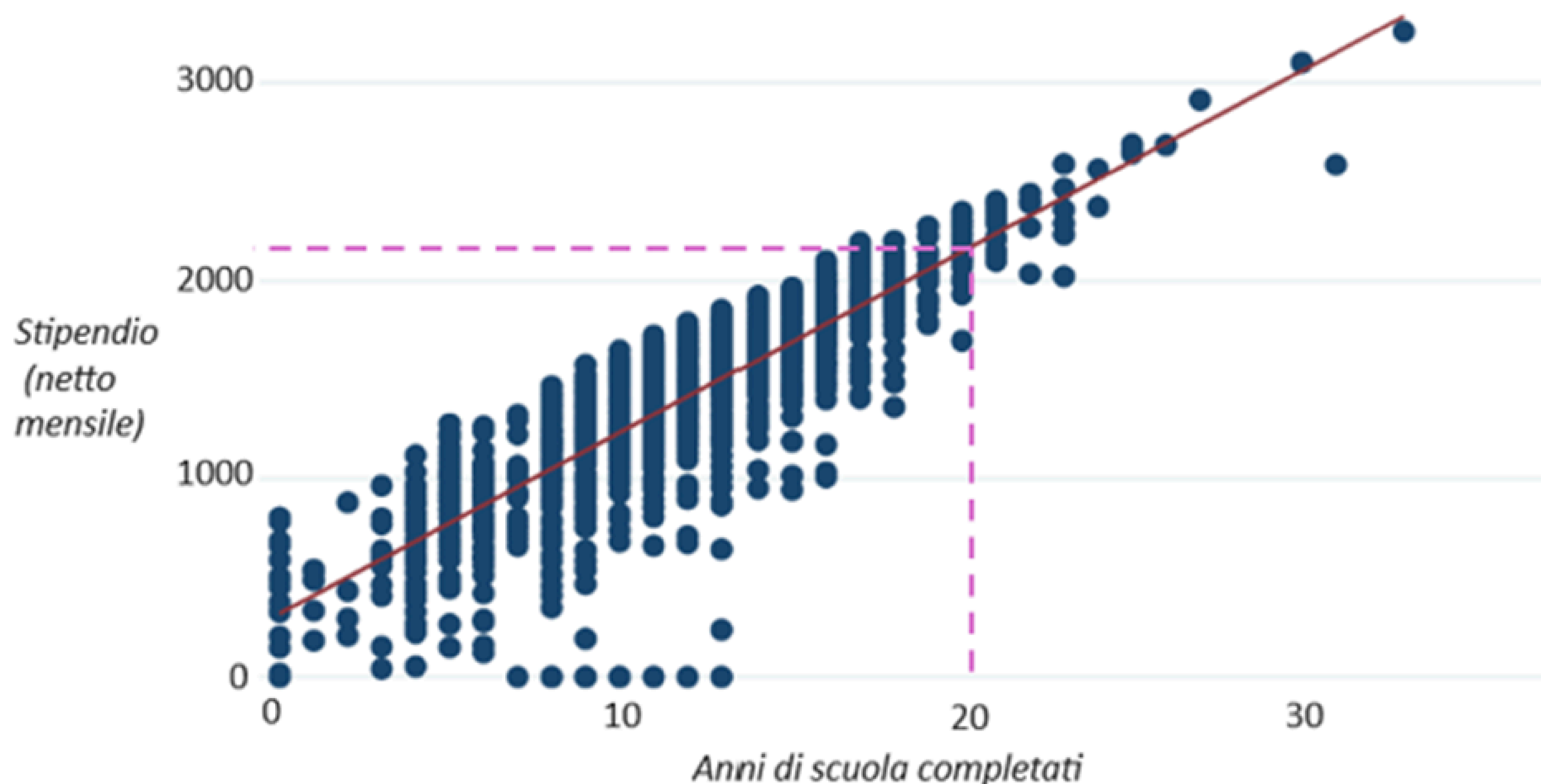
3

RETTA DI REGRESSIONE E DISPERSIONE

I puntini blu indicano le nostre osservazioni. La retta di regressione (in rosso) suggerisce un andamento tra le due variabili: all'aumentare degli anni di scuola completati da un individuo, aumenta lo stipendio netto mensile.

L'inclinazione della curva (beta) ci indica di quanto aumenta lo stipendio aggiungendo un anno di scuola completato in più. In media, gli individui con 20 anni di istruzione hanno uno stipendio netto mensile di circa 2150 (tirando una linea verticale da 20, fino a dove si incontra la retta di regressione, e poi tirando una retta in orizzontale).

La retta ci permette di stimare la relazione tra due variabili anche quando non abbiamo osservazioni (per esempio: possiamo stimare lo stipendio netto mensile di un individuo con 40 anni di scuola completati anche se non abbiamo individui con quelle caratteristiche nel nostro campione).



4

RETTA DI REGRESSIONE E CORRELAZIONE

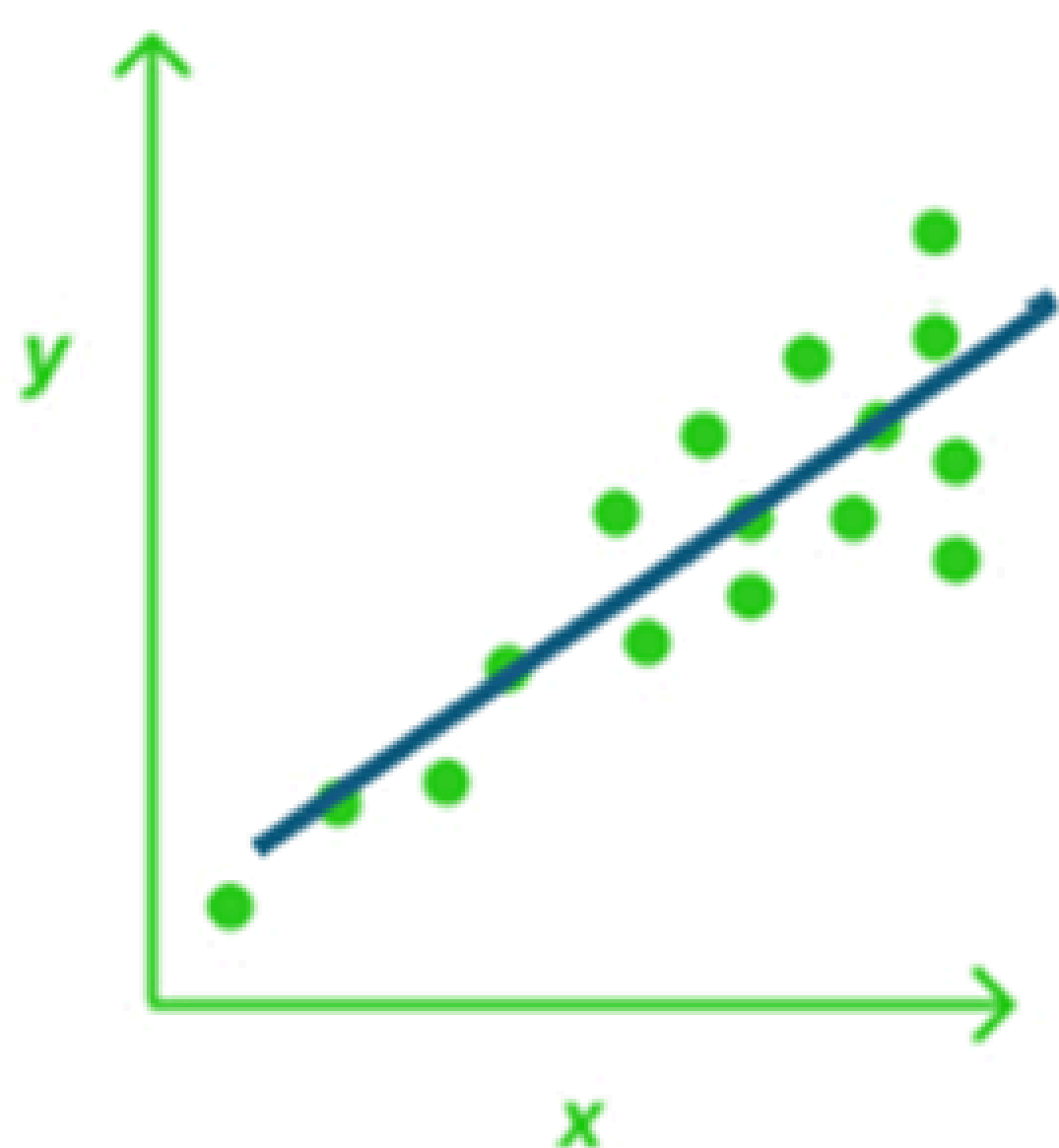
Esistono diversi tipi di relazioni tra variabili che si evidenziano con le rette di regressioni e i relativi grafici.

Se la relazione (correlazione) è positiva, significa che all'aumentare del valore della variabile indipendente X aumenta anche il valore della variabile dipendente Y (per esempio: ore di studio (X) e voto (Y)).

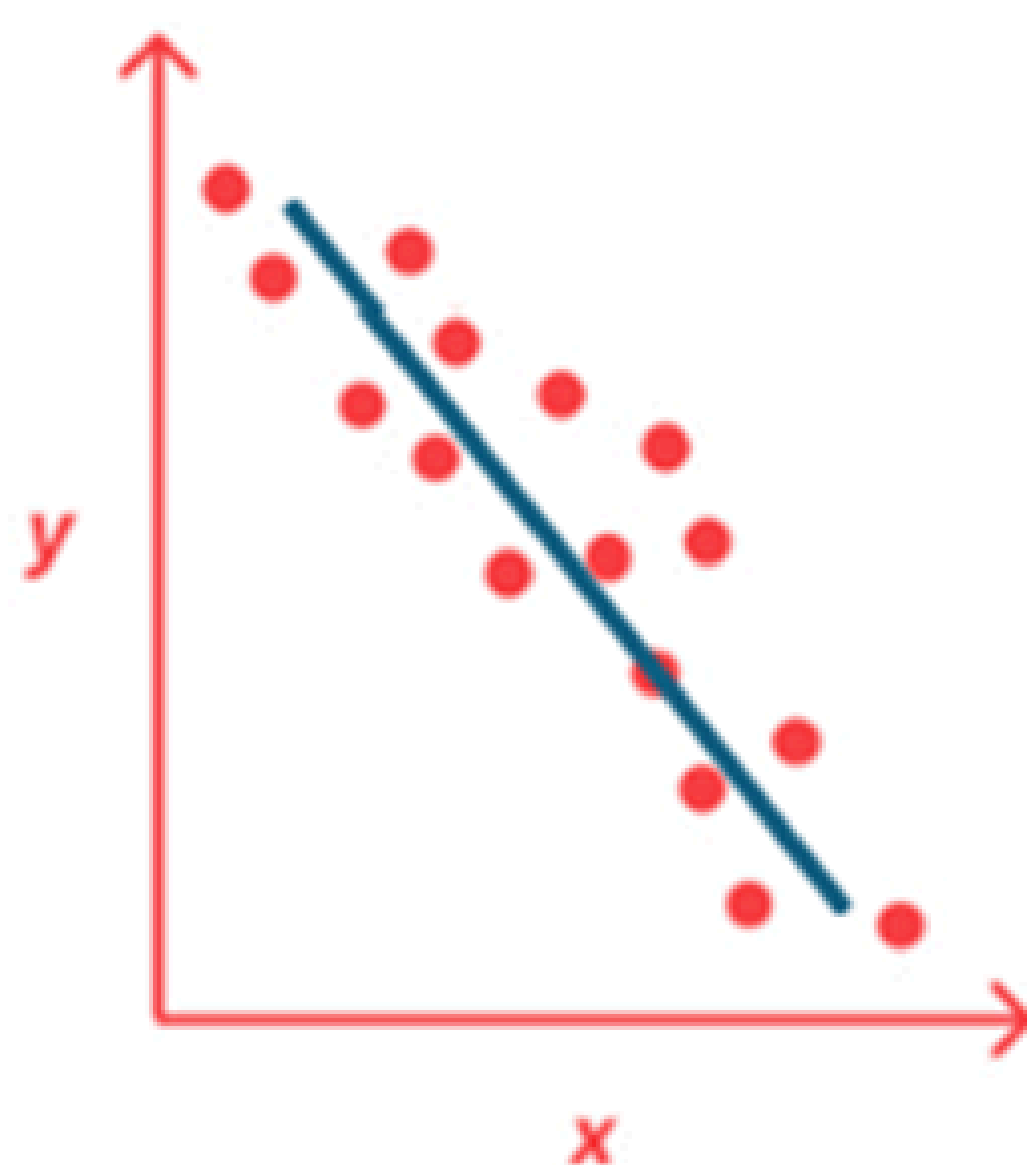
Se la relazione (correlazione) è negativa, significa che all'aumentare del valore della variabile indipendente X diminuisce il valore della variabile dipendente Y (per esempio: ore di assenza (X) e livelli di motivazione (Y)). Se la relazione è neutra, significa che non possiamo stabilire una correlazione tra le nostre due variabili, ovvero non sono correlate.

Relazione Positiva

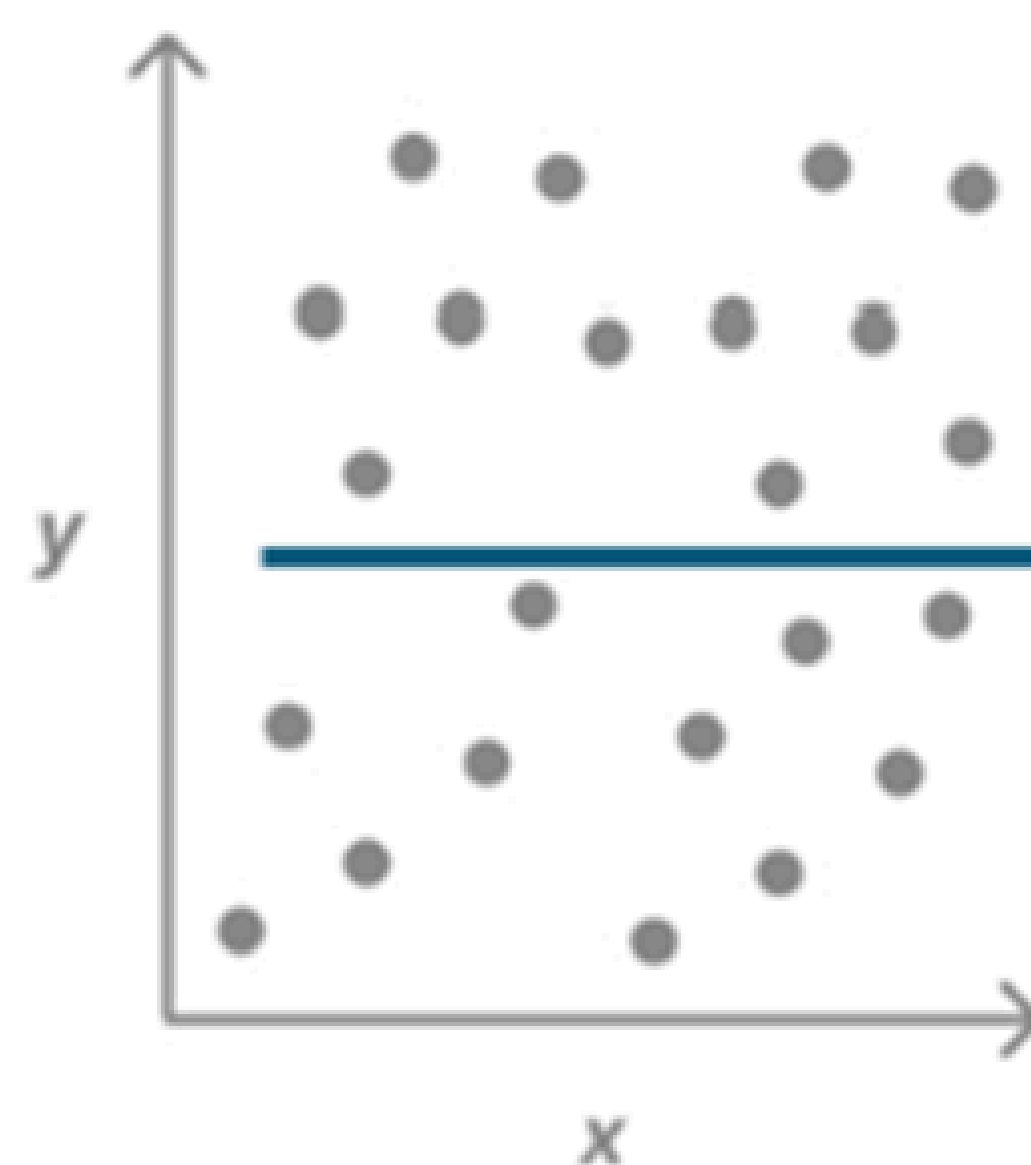
Se aumenta X , aumenta Y

Relazione Negativa

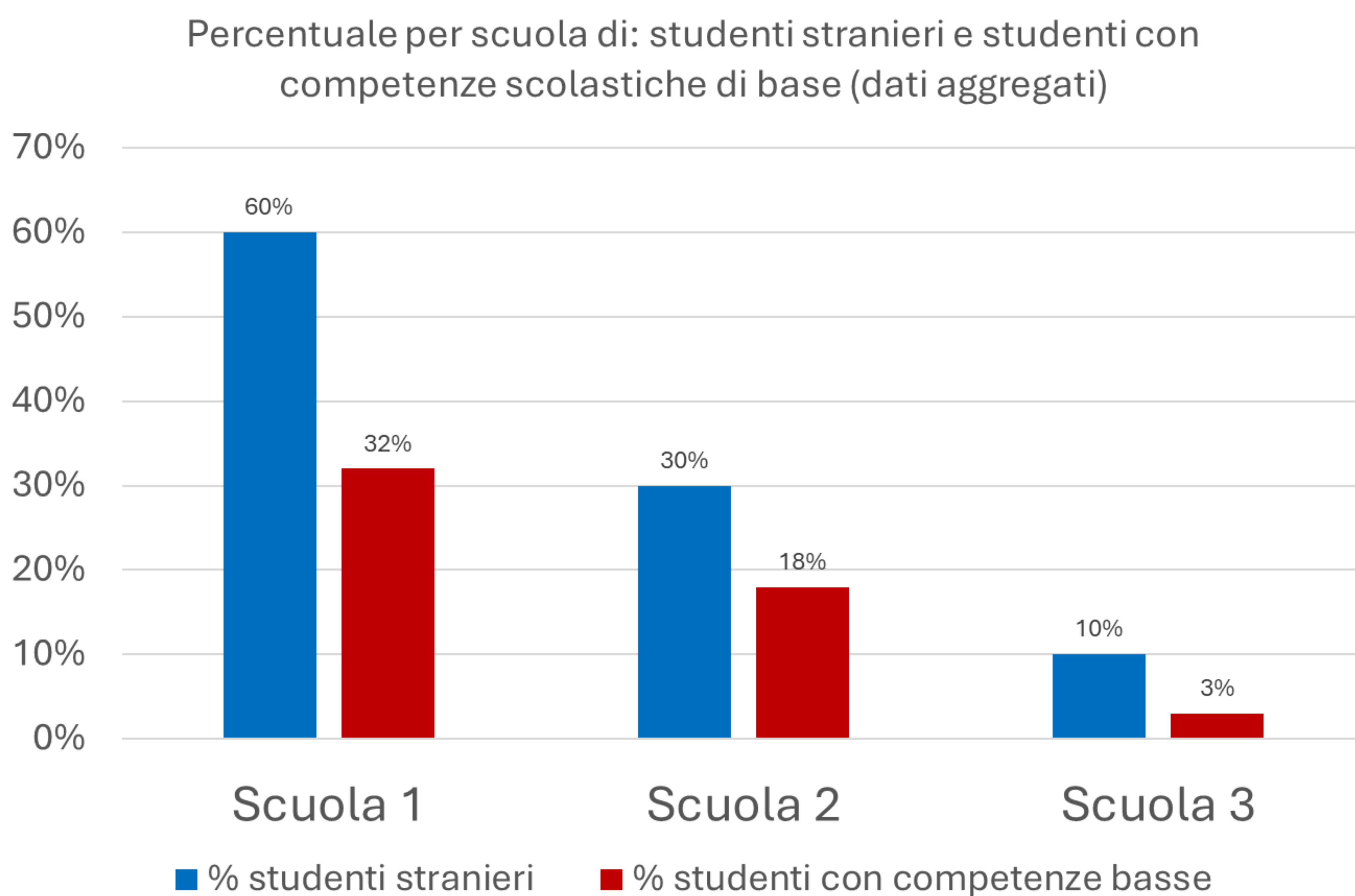
Se aumenta X , diminuisce Y

Relazione Neutra

Se aumenta X ,
 Y aumenta o diminuisce

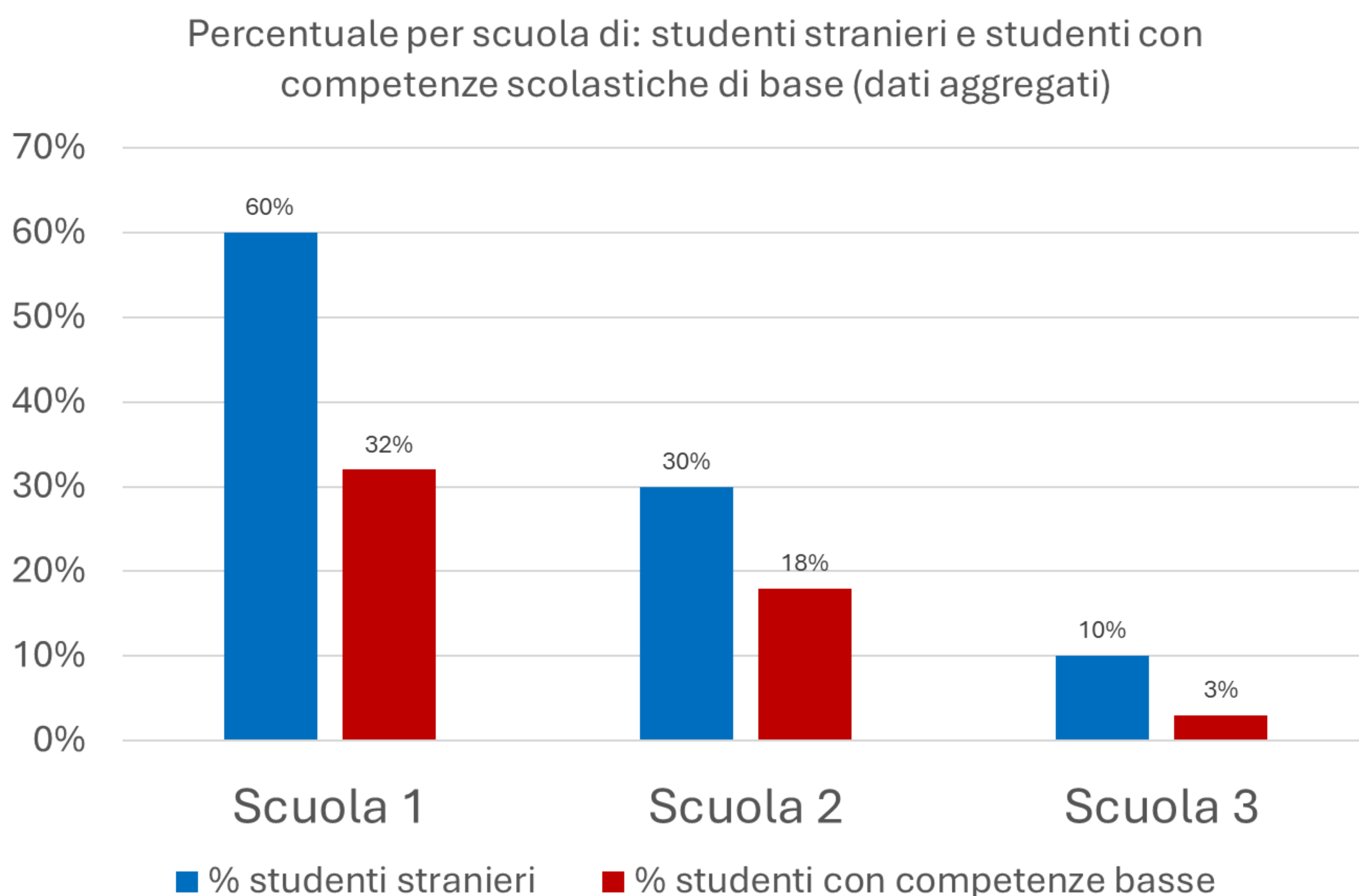


LEGGIAMO IL GRAFICO



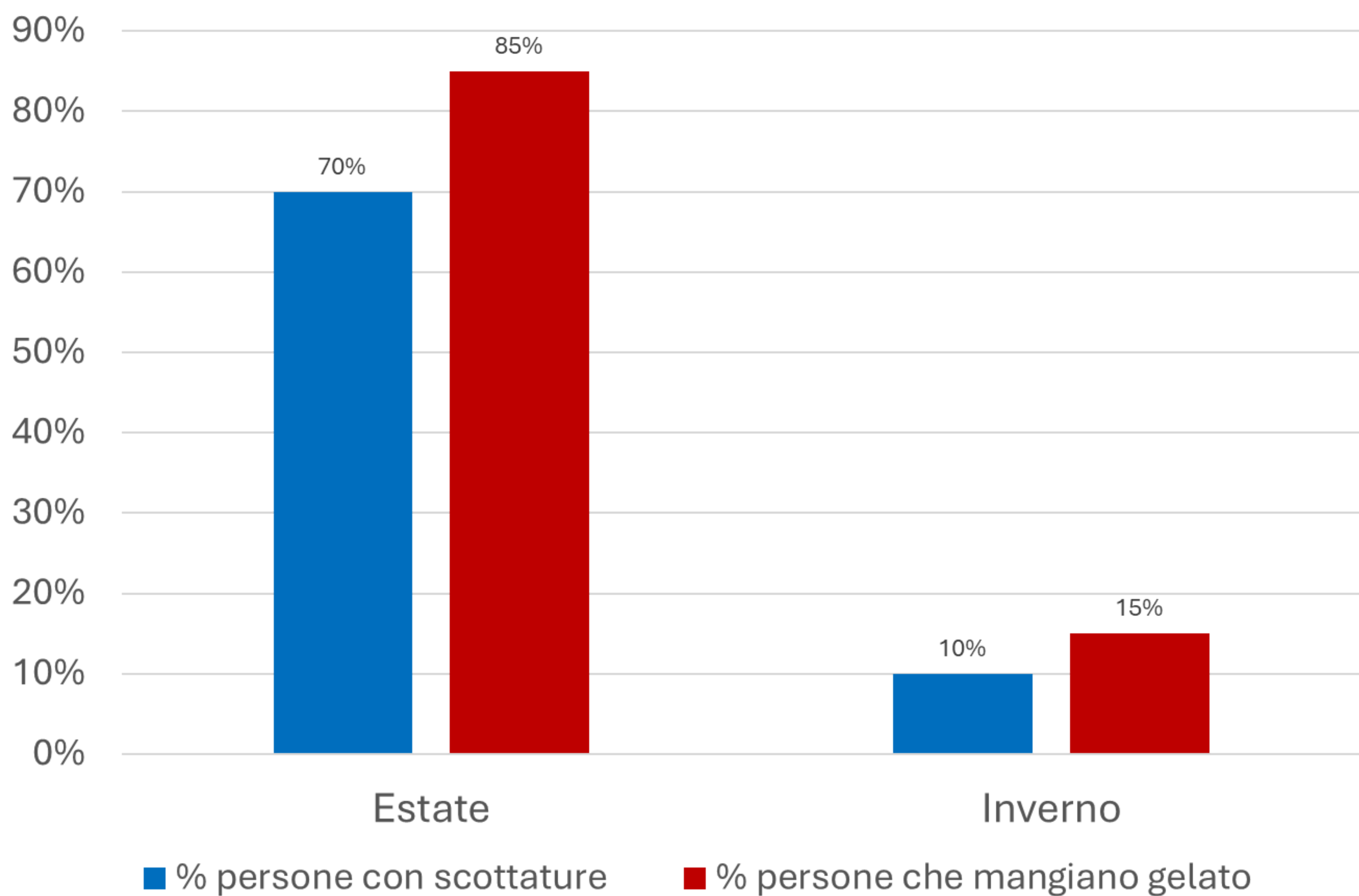
- Nella scuola 1, il 60 % di studenti è di origine straniera. Nella scuola 1, il 32% degli studenti registra delle competenze scolastiche di base.
- Nella scuola 2, il 30% degli studenti è di origine straniera, e il 18% registra competenze scolastiche di base (ecc).
- Per identificare una relazione tra due variabili, servono test statistici precisi (p-value).
- A livello aggregato (di scuola) i test ci suggeriscono una associazione positiva tra % di studenti stranieri e % di studenti con competenze scolastiche di base (all'aumentare di uno, è probabile che aumenti anche l'altro).

COSA NON CI DICE QUESTO GRAFICO



- Gli studenti stranieri sono quelli con più competenze scolastiche di base
NO! Non abbiamo i dati micro (a livello di studente) quindi non possiamo sapere le caratteristiche degli studenti all'interno della scuola che registrano competenze di base (sono di origine straniera o sono italiani?)
- Avere più studenti stranieri in una scuola causa maggiore ritardo scolastico generale
NO! CORRELAZIONE non significa CAUSAZIONE

CORRELAZIONE NON SIGNIFICA CAUSAZIONE



- **C'è una correlazione positiva tra persone che mangiano gelato e persone con le scottature.**
- **Questo NON significa che mangiare il gelato causi scottature!**
- **Tornando all'esempio precedente, non possiamo dedurre dal grafico che essere studenti stranieri causi maggiori ritardi scolastici.**
- **Ci sono altre variabili che possono «spiegare» la correlazione positiva. Esempio: contesto geografico e/o socioeconomico delle scuole.**
- **Nelle scienze sociali, identificare una causazione prevede modelli statistici complessi. La maggior parte degli studi identifica correlazioni tra due o più variabili.**