

# Regressione lineare

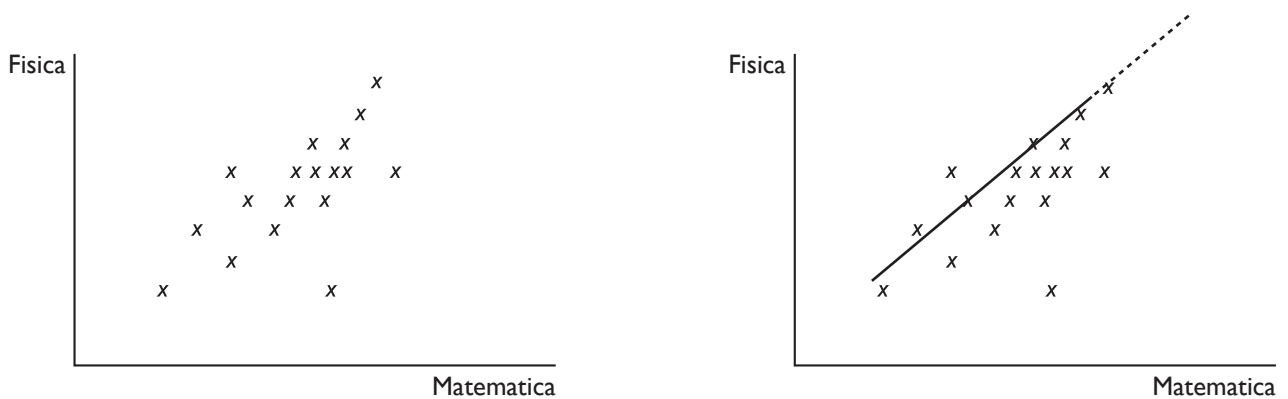
Scuola media inferiore

## Obiettivo

Esaminare il modo in cui una serie bivariata (a due variabili) di dati può essere utilizzata per stimare il valore di una delle due variabili per un dato valore dell'altra.

## Spiegazione dell'attività

Per molte distribuzioni bivariate, sembra esserci un collegamento tra i due insiemi di dati. Per esempio, se raccogliamo i voti ottenuti in un esame di scienze da una classe di studenti e raccogliamo i voti conseguiti dagli stessi studenti in un esame di matematica, possiamo osservare che uno studente con buoni voti in scienze potrebbe avere buoni voti anche in matematica. Tra questi voti sembra esserci una *correlazione positiva*. Se riportiamo i voti di ogni studente in un grafico, la linea che si adatta meglio all'intervallo di variabilità dei voti può essere utilizzata per stimare il valore di un voto rispetto a un altro valore dato. La tracciatura dei voti d'esame produce un *diagramma di dispersione*.



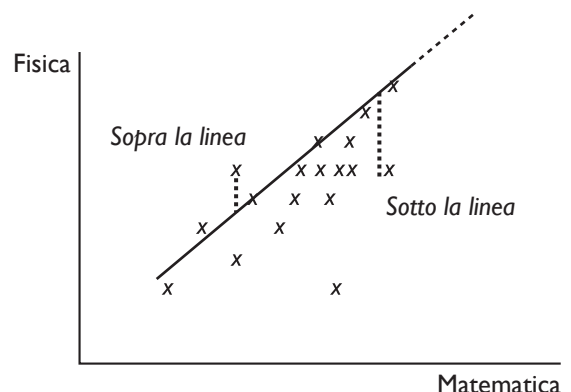
La linea è denominata linea di regressione lineare e per trovarne l'equazione possiamo utilizzare la calcolatrice.

Se prendiamo in considerazione il precedente esempio della linea tracciata nel diagramma di dispersione, per una retta si avrà una ben nota equazione:

$$y = a + b x$$

La posizione della linea nel diagramma di dispersione può essere determinata tracciando la linea in modo tale che, quando eleviamo al quadrato *tutte* le differenze individuali al di sopra della linea e *tutte* le differenze individuali al di sotto della linea, la somma di tutti questi quadrati sia pari al valore minimo.

Questo è noto come il **metodo dei minimi quadrati** e la linea in questione è detta **linea di regressione di y su x**. È questo il processo utilizzato dalla calcolatrice.

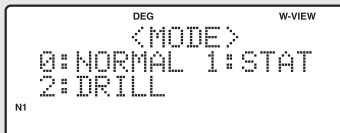


Quando una serie di dati a due variabili è stata immessa correttamente, la calcolatrice può essere utilizzata per trovare i valori di **a** e **b**, allo scopo di formulare l'equazione della linea di regressione con la quale possiamo effettuare delle stime della variabile y, conoscendo un valore dell'altra variabile x.

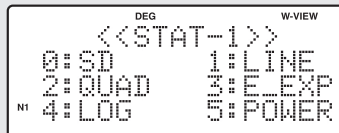
•••••••••• **Uso della calcolatrice** ••••••••••

Impostare la calcolatrice in modalità Statistica.

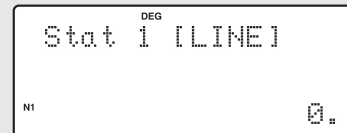
Tasto MODALITÀ



Tasto numerico 1



Tasto numerico 1

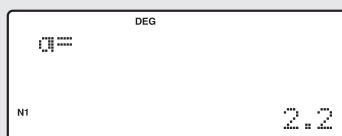


La calcolatrice è ora impostata per l'immissione di dati a due variabili. Ogni coppia  $p, q$  viene inserita come  $p$  [tasto  $x, y$ ]  $q$  [tasto DATA]. Per esempio, se la coppia bivariata 0, 2 fosse il primo elemento di un insieme di dati, apparirebbe sul display della calcolatrice come a fianco indicato.

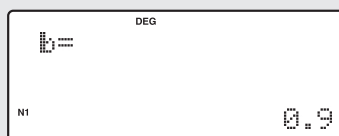
Immettere i dati bivariati riportati nella seguente tabella (le schermate relative all'intero processo sono riportate a fianco):

$p$	0	1	2	3	4
$q$	2	3	5	4	6

Una volta immessi i dati, i valori di  $a$  e  $b$  dell'equazione relativa alla linea di regressione di  $y$  su  $x$ ,  $y = a + b x$ , possono essere trovati mediante l'uso della calcolatrice. Il tasto ALPHA identifica  $a$  e  $b$  come funzioni supplementari sui tasti "parentesi".



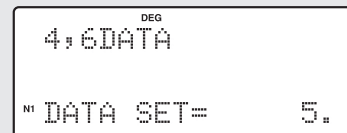
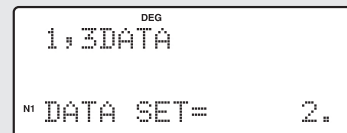
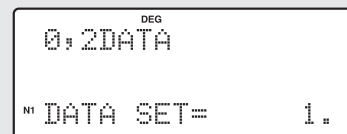
ALPHA  $a$  =



ALPHA  $b$  =

La linea di regressione di  $y$  su  $x$  è quindi  $y = 2.2 + 0.9 x$

Le stime del valore di  $y$  possono essere trovate immettendo dati valori di  $x$  nell'equazione.



•••••••• **Punti di discussione per gli studenti** ••••••••

Poiché la linea di regressione di  $y$  su  $x$  non è perfetta per tutti i punti del diagramma di dispersione, viene spesso descritta come la linea migliore. Ne deriva che, per qualsiasi valore dato di  $x$ , è possibile ottenere soltanto delle stime del valore  $y$ .

Il concetto di connessione tra due variabili implica i concetti di correlazione positiva (entrambe le variabili aumentano o diminuiscono contemporaneamente) e di correlazione negativa (una variabile aumenta e l'altra diminuisce).

Un coefficiente di correlazione è un modo per tentare di misurare numericamente tale connessione.