

## Anche le rette raccontano

**Livello scolastico:** 2° biennio

<b>Abilità interessate</b>	<b>Conoscenze</b>	<b>Nuclei coinvolti</b>	<b>Collegamenti esterni</b>
Identificare situazioni che richiedono di rilevare lo stesso carattere su una unità statistica formata da due elementi, o 2 caratteri diversi sulla stessa unità statistica.	Concetto e significato di modello: correlazione e regressione.	<u>Dati e previsioni</u> Spazio e figure Argomentare, congetturare, dimostrare Risolvere e porsi problemi Laboratorio di matematica	Organizzazione delle attività

### Contesto

Extramatematico, sociale.

Il contesto è di tipo matematico ed extramatematico; in particolare, per il contesto matematico, si pone in evidenza l'ambito statistico (rappresentazioni grafiche; correlazione e regressione).

L'attività riguarda lo studio di una distribuzione doppia rispetto a due caratteri entrambi quantitativi e può essere introdotta, nella forma che qui viene proposta, nel secondo biennio quando gli studenti hanno acquisito sicurezza sulle principali abilità relative all'uso del foglio elettronico. Si richiede l'uso dei filtri in una tabella, la padronanza delle diverse tipologie di indirizzamento e l'uso di alcune funzioni statistiche. Se non si dispone del foglio di calcolo, le proposte indicate di seguito sono ugualmente valide, ma richiedono più tempo per l'elaborazione.

### Descrizione dell'attività

L'esempio presenta l'esito finale di una classe prima in tre materie. La tabella è stata tratta dal database dell'unità "A proposito di valutazione scolastica" con l'uso del filtro appropriato.

Alunno	Matematica	Fisica	Inglese	Alunno	Matematica	Fisica	Inglese
	x	y	z		x	y	z
1	5	4	5	14	8	8	8
2	5	6	4	15	9	8	8
3	4	5	4	16	3	4	6
4	6	6	5	17	5	6	6
5	8	8	8	18	6	7	8
6	6	6	6	19	6	6	8
7	6	6	6	20	6	7	8
8	7	8	7	21	6	6	8
9	9	8	7	22	6	6	6
10	6	6	7	23	5	6	7
11	8	7	7	24	6	7	6
12	5	5	6	25	7	8	6
13	8	8	8	26	5	5	5

*Tabella 1*

Prima fase

L'insegnante propone all'osservazione della classe la Tabella 1 e formula alcune domande. Chi è bravo in matematica lo è anche in fisica o in inglese? Come si può stabilire se uno studente è bravo in una determinata materia? In questa classe chi ha 7 in matematica è "bravo"?

L'insegnante propone la Figura 1 e stimola gli studenti a riflettere sul grafico in essa contenuto.

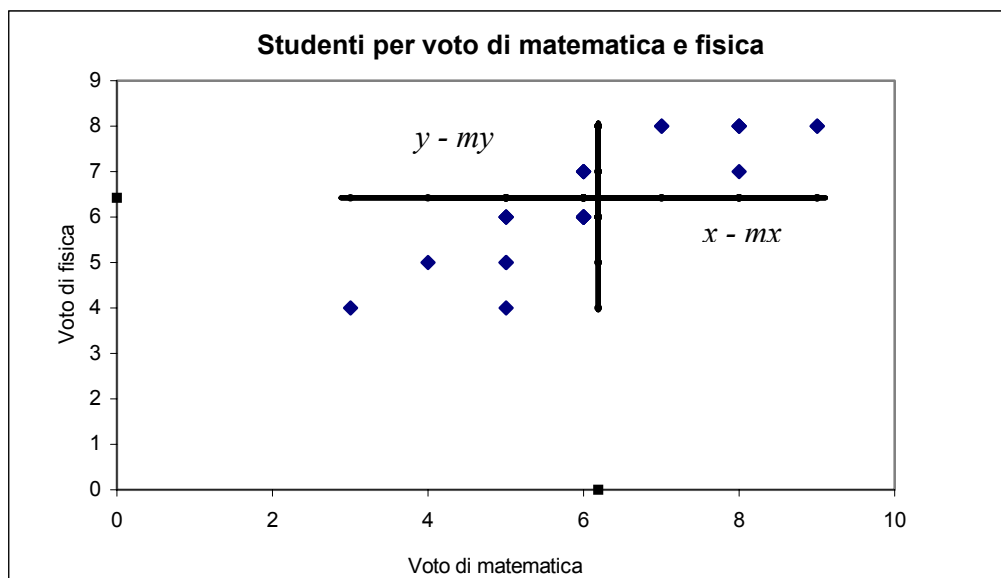


Figura 1

Come mai i punti sono meno di 26 (numero degli studenti)? La dislocazione dei punti sul piano cartesiano dà una informazione su un eventuale legame tra il voto di matematica ( $x$ ) e il voto di fisica ( $y$ )? Perché sono stati rappresentati gli scarti dalla media aritmetica ( $x - mx$ ) e ( $y - my$ )? Che informazioni aggiuntive offrono?

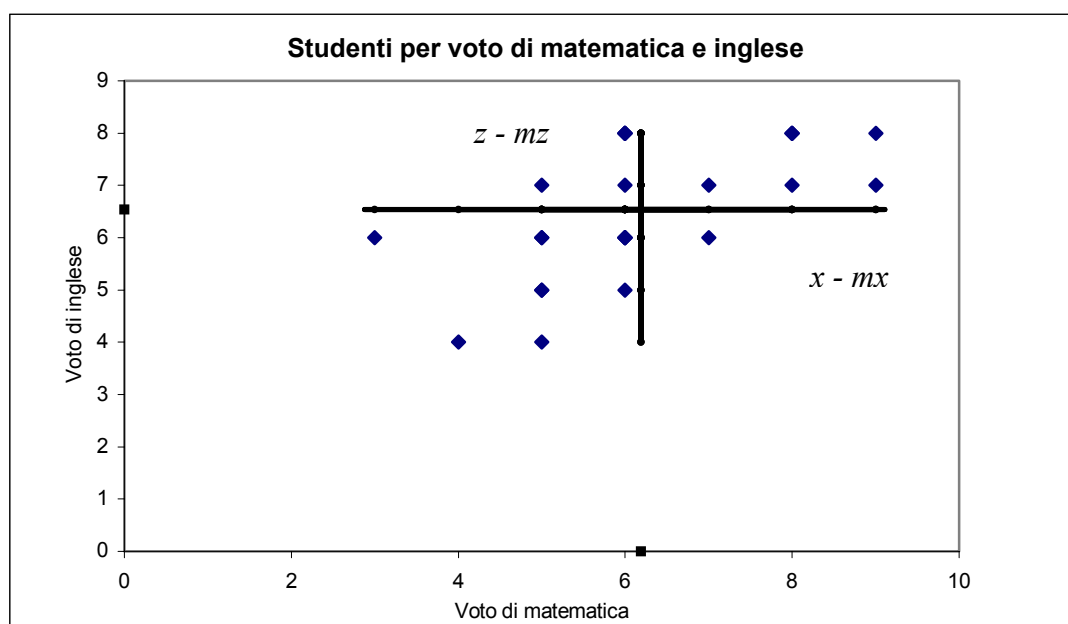


Figura 2

Successivamente l'insegnante stimola gli studenti a riflettere sul grafico di Figura 1.

Gli studenti più bravi in matematica sono più bravi anche in fisica? E' vero anche il viceversa? Ci sono studenti più bravi in matematica, ma meno bravi in fisica? E' vero anche il viceversa? Rispetto al sistema traslato nel baricentro  $[(x - mx); (y - my)]$  quali segni hanno le coordinate di queste unità statistiche? L'insegnante sottolinea che a questo punto è possibile sfruttare la nota relazione tra i segni di due numeri relativi per cui il prodotto tra due concordi è positivo e il prodotto tra due discordi è negativo.

Rispetto al sistema degli scarti si può dunque dire che se i punti della nuvola si dispongono con maggiore frequenza nel primo e nel terzo quadrante fra le due variabili vi è "concordanza", mentre in caso contrario vi è "discordanza". In Figura 1 è evidente la concordanza, ossia al crescere del voto in matematica cresce mediamente anche il voto in fisica. A queste prime valutazioni soggettive è tuttavia indispensabile sostituire una misura oggettiva, che si fonda sul prodotto degli scarti che competono a ciascuna unità.

Per potere confrontare misure ottenute su distribuzioni diverse è generalmente indispensabile che esse varino fra 0 e 1. In questo particolare ambito concettuale però è necessario che si tenga conto che la correlazione ha un doppio aspetto: concordanza o discordanza. Questa duplicità si esprime con i segni: + esprime la concordanza, - esprime la discordanza, dunque ciò che si vuole è un indice che assuma valori fra -1 e +1. Se assume valori positivi è un indice di concordanza, se assume valori negativi è un indice di discordanza. Il valore 0 è indicatore di una situazione da esaminare con attenzione poiché si può realizzare o in ipotesi di indipendenza, connessione nulla, tra i caratteri oppure quando i punti, nel loro insieme, non "danno l'idea" di un andamento lineare.

Il "coefficiente di correlazione lineare di Bravais-Pearson" è la soluzione a questo problema concettuale. Esso ha la seguente espressione:

$$r = \frac{\sum_{i=1}^n (x_i - mx)(y_i - my)}{\sqrt{\sum_{i=1}^n (x_i - mx)^2 \cdot \sum_{i=1}^n (y_i - my)^2}}$$

Da chi dipende il segno di  $r$ ?  $r$  ha un'unità di misura? Perché?

Il coefficiente di correlazione lineare di Bravais-Pearson può essere agevolmente determinato utilizzando le potenzialità del foglio elettronico, evitando che gli studenti si concentrino eccessivamente sull'aspetto puramente di calcolo tralasciando quello interpretativo.

Alunno	Matematica	Fisica					
	$x$	$y$	$x - mx$	$y - my$	$(x - mx)^2$	$(y - my)^2$	$(x - mx)(y - my)$
1	5	4	-1,192	-2,423	1,422	5,871	2,88905
2	5	6	-1,192	-0,423	1,422	0,179	0,50444
3	4	5	-2,192	-1,423	4,806	2,025	3,11982
4	6	6	-0,192	-0,423	0,037	0,179	0,08136
5	8	8	1,808	1,577	3,268	2,487	2,85059
...	...	...	...	...	...	...	...
26	5	5	-1,192	-1,423	1,422	2,025	1,69675
Totali	161	167	0	0	54,0385	38,3462	39,8846
Media ar.	6,192	6,423					

Tabella 2

La Tabella 2 riporta soltanto i calcoli relativi alle prime cinque unità statistiche su 26 e fa riferimento al "Voto in matematica" e il "Voto in fisica" ed utilizza il foglio di calcolo Excel.

Perché la somma della 4° e 5° colonna è uguale a 0? Quale significato si può attribuire al segno della somma dell'ultima colonna? Un valore positivo come quello ottenuto indica che tutti gli scarti dalla media aritmetica sono positivi?

Il coefficiente di correlazione lineare  $r$  calcolato rispetto ai voti di matematica e fisica è:

$$r = \frac{39,88}{\sqrt{54,04 \cdot 38,35}} = 0,8762$$

Il valore ottenuto è coerente con le osservazioni fatte sulla Figura 1? Sapendo che  $r = 1$  quando i punti della nuvola sono allineati su una retta crescente, l'insegnante chiede: dai dati a disposizione ci si poteva attendere un valore di  $r$  così alto?

E' un utile esercizio di verifica ricorrere alla funzione statistica messa a disposizione dal foglio elettronico Excel che consente di calcolare in modo automatico il coefficiente di correlazione lineare. Si indica di seguito il percorso da seguire per ottenerne il valore.

Attivare il pulsante "Incolla funzione" per aprire la finestra riportata in Figura 3:

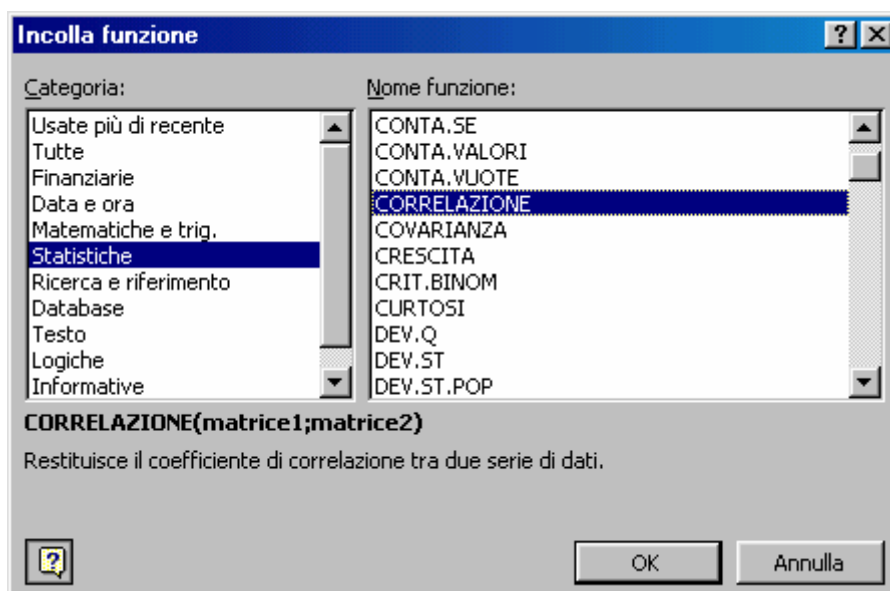


Figura 3

Dopo aver confermato con OK si ottiene la finestra della Figura 4:

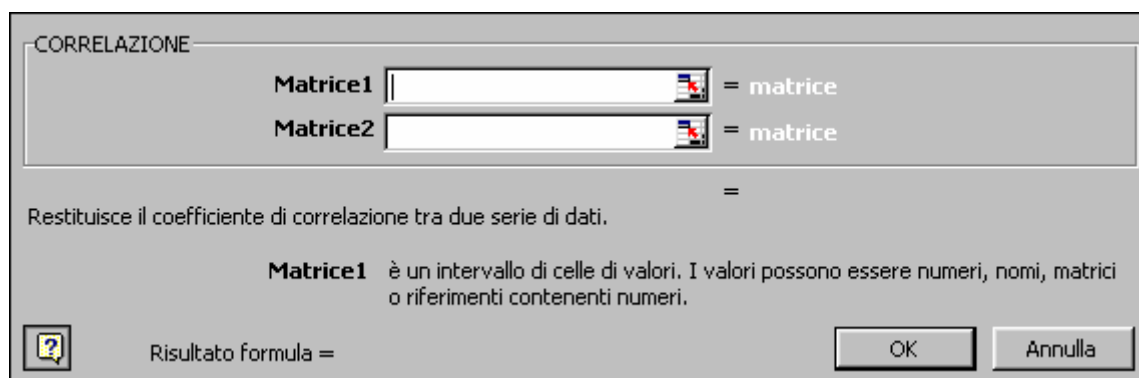


Figura 4

“Matrice1” è un intervallo di celle di valori e, in questo caso, l’intervallo dei valori dei voti di matematica, “Matrice2” è il secondo intervallo di celle di valori e, sempre in questo caso, l’intervallo dei valori dei voti di fisica.

L’indicazione riportata a fianco di Matrice 1 va intesa, nel calcolo del coefficiente di correlazione lineare, come limitata ai valori numerici.

La chiusura della finestra fornisce il valore del coefficiente cercato.

Il risultato ottenuto con la funzione “Correlazione( )” del foglio elettronico coincide con quello ottenuto dai calcoli effettuati seguendo tutti i passaggi?

Osservando la Figura 2, che si riferisce ai voti di matematica ed inglese, si può pensare che la somma dei prodotti degli scarti dalla media aritmetica sia positiva anche in questo caso?

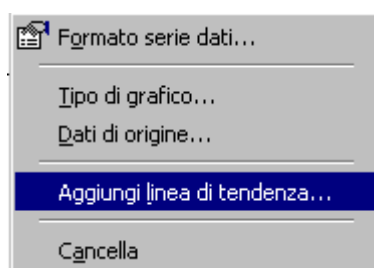
L’insegnante propone come esercizio il calcolo del coefficiente  $r$  per la distribuzione doppia dei caratteri “Voto in matematica” e “Voto in inglese”.

### Seconda fase

La forma della nuvola dei punti delle figure precedenti e l’alto valore di  $r$ , nel caso esaminato, giustificano la ricerca di una retta che interpoli l’insieme dei punti che rappresentano il fenomeno, passando il più vicino ad essi, secondo il “metodo dei minimi quadrati”.

La determinazione di questa funzione può essere fatta, in modo agevole, usando le funzioni del foglio elettronico Excel, come evidenziato di seguito.

Occorre selezionare un punto qualsiasi sul grafico e, cliccando con il pulsante destro, far apparire il menu evidenziato nella Figura 5:



*Figura 5*

La voce evidenziata in Figura 5 fa apparire la finestra “Aggiungi linea di tendenza”. La scheda “Tipo” consente di scegliere una funzione interpolatrice, lineare nel caso in esame<sup>1</sup>, mentre la scheda “Opzioni” consente, selezionando le voci spuntate come in Figura 6, di tracciare sul grafico la retta interpolatrice, secondo il metodo dei minimi quadrati, come si vede in Figura 7.

<sup>1</sup> Occorre porre attenzione al fatto che, pur scegliendo come funzione interpolatrice una funzione non lineare, il problema della ricerca dei parametri risolto con il metodo dei minimi quadrati è pur sempre lineare a patto che l’interpolatrice sia di tipo polinomiale o ad esso riconducibile. Si parla in tal caso di modello lineare nei parametri.

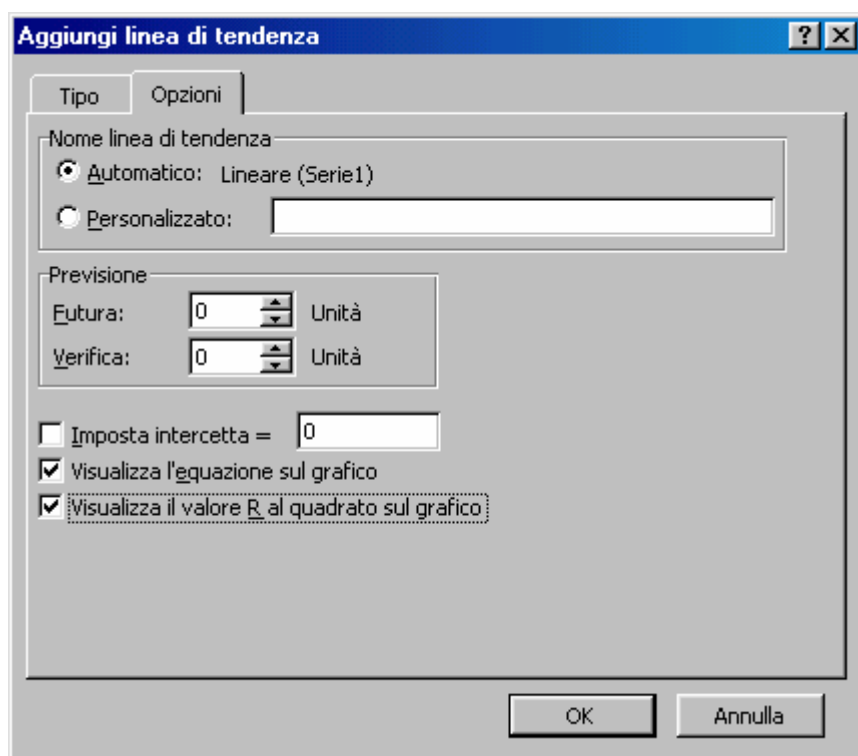


Figura 6

Nella Figura 7, oltre al grafico della retta interpolatrice, il foglio elettronico fornisce anche la sua equazione e l'“Indice di determinazione lineare”  $R^2$ .

$R^2$  è un indice che può assumere valori compresi tra 0 ed 1. Esso vale 1 quando tutti i punti osservati appartengono ad una retta, crescente o decrescente. In tale caso quella retta esprime la dipendenza lineare che esiste tra  $X$  ed  $Y$  in base alla quale il valore di  $Y$  è determinato quando si conosce quello di  $X$ .  $R^2$  vale 0 sia quando esiste indipendenza fra i caratteri, sia quando l'andamento della nuvola dei punti non consente di proporre l'interpolazione con una retta. In questo caso la retta interpolatrice ha equazione  $y = 0$  e nota la variabile  $X$  non può dirsi nulla sul valore che assume l'altra variabile.

L'insegnante stimola l'attenzione della classe, ponendo una serie di domande sulla Figura 7.

Perché la retta passa per l'origine? C'è una relazione tra il coefficiente  $r$  di correlazione lineare di Bravais–Pearson e il coefficiente  $R^2$  di determinazione lineare?

L'insegnante guida gli studenti ad osservare che, per una distribuzione doppia,  $R^2$  altro non è che il quadrato di  $r$ .

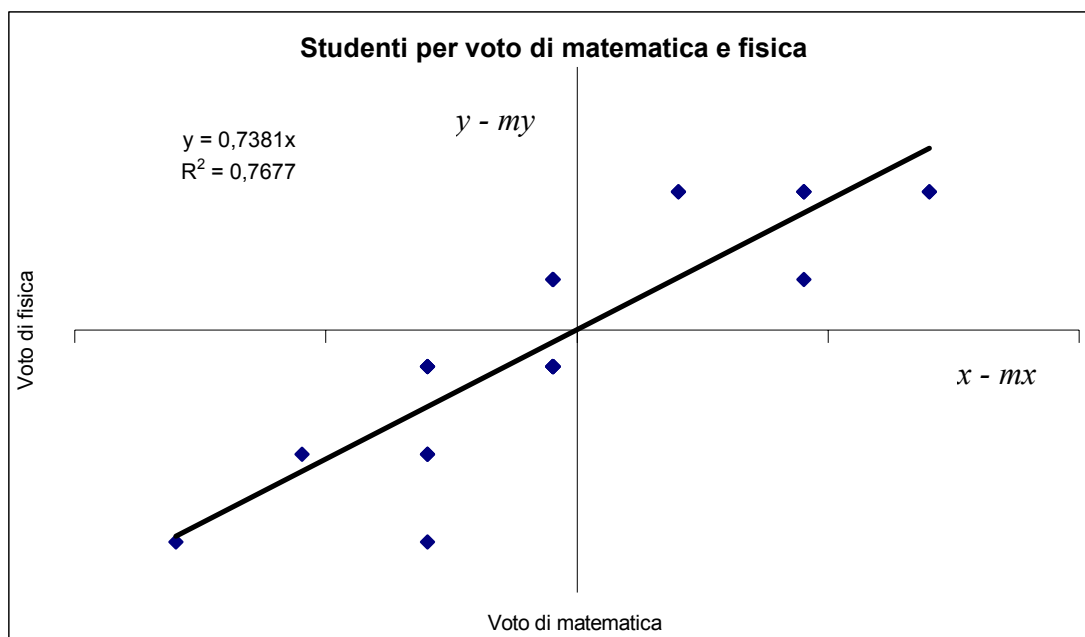


Figura 7

L'equazione della retta interpolatrice fornita dal foglio elettronico è coerente con quanto ci si attende, poiché manca il termine noto. Come si giustifica il valore del coefficiente angolare della retta? Nell'elaborazione dei dati del foglio elettronico usati in precedenza è possibile trovare quelli necessari al calcolo del "Coefficiente angolare ( $b$ ) della retta interpolatrice"? L'insegnante mostra agli studenti la formula di  $b$  che risolve il problema<sup>2</sup>. Tale formula, applicata alla Tabella 2, consente arrivare alla soluzione anche quando non si ha a disposizione il computer.

Gli stessi risultati si possono ottenere utilizzando alcune funzioni di tipo statistico del foglio elettronico:

PENDENZA

**Y\_nota**  = matrice

**X\_nota**  = matrice

=

Restituisce la pendenza della retta di regressione lineare fra le coordinate note.

**X\_nota** è l'insieme dei valori indipendenti e può essere costituito da numeri, nomi, matrici o riferimenti contenenti numeri.

Risultato formula =

OK Annulla

Figura 8

La funzione "Pendenza()" restituisce il coefficiente angolare della retta di regressione lineare.

<sup>2</sup> La formula di  $b$  è:

$$b = \frac{\sum_{i=1}^n (x_i - mx)(y_i - my)}{\sum_{i=1}^n (x_i - mx)^2}$$

Qual è l'interpretazione statistica del coefficiente di regressione? Si può dire che ad ogni incremento unitario del voto di matematica ( $X$ ) il voto in fisica ( $Y$ ) aumenta di  $b$ ? In quale relazione stanno tra loro  $r$  e  $b$ ?

L'ordinata all'origine in questo caso vale 0, riportando il sistema di assi cartesiani nell'origine, come cambia il termine noto  $a$ ? Quale interpretazione ha il valore trovato?

Quando il problema richiede di calcolare anche l'intercetta  $a$ , il foglio elettronico Excel mette a disposizione la funzione statistica "Intercetta( )" che ha una sintassi del tutto analoga alla funzione di Figura 8.